Advances in Intelligent Systems and Computing 387

Thi Thi Zin
Jerry Chun-Wei Lin
Jeng-Shyang Pan
Pyke Tin
Mitsuhiro Yokota *Editors*

Genetic and Evolutionary Computing

Proceedings of the Ninth International Conference on Genetic and Evolutionary Computing, August 26-28, 2015, Yangon, Myanmar – Volume I

Springer

Advances in Intelligent Systems and Computing

Volume 387

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu
Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es
Hani Hagras, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk
László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu
Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu
Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw
Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au
Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org
Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br
Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl
Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at http://www.springer.com/series/11156

Thi Thi Zin · Jerry Chun-Wei Lin
Jeng-Shyang Pan · Pyke Tin
Mitsuhiro Yokota
Editors

Genetic and Evolutionary Computing

Proceedings of the Ninth International Conference on Genetic and Evolutionary Computing, August 26–28, 2015, Yangon, Myanmar – Volume I

Springer

Editors
Thi Thi Zin
Faculty of Engineering
University of Miyazaki
Miyazaki
Japan

Jerry Chun-Wei Lin
School of Computer Science and Technology
Harbin Institute of Technology Shenzhen Graduate School
Shenzhen
China

Jeng-Shyang Pan
College of Information Science and Engineering
Fujian University of Technology
Fuzhou
China

Pyke Tin
Faculty of Engineering
University of Miyazaki
Miyazaki
Japan

Mitsuhiro Yokota
Faculty of Engineering
University of Miyazaki
Miyazaki
Japan

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-23203-4 ISBN 978-3-319-23204-1 (eBook)
DOI 10.1007/978-3-319-23204-1

Library of Congress Control Number: 2015946740

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Preface

This volume composes the proceedings of the Ninth International Conference on Genetic and Evolutionary Computing (ICGEC 2015), which was hosted by University of Computer Studies, Yangon and was held in Yangon, Myanmar on 26–28, August, 2015. ICGEC 2015 was technically co-sponsored by Springer, Ministry of Science and Technology, Myanmar, University of Computer Studies, Yangon, University of Miyazaki in Japan, Kaohsiung University of Applied Science in Taiwan, Fujian University of Technology in China and VSB-Technical University of Ostrava. It aimed to bring together researchers, engineers, and policymakers to discuss the related techniques, to exchange research ideas, and to make friends.

93 excellent papers were accepted for the final proceeding. Three plenary talks were kindly offered by: Professor Chin-Chen Chang (IEEE Fellow, IET Fellow, Feng Chia University, Taiwan), Professor Yutaka Ishibashi (IEICE Fellow, Nagoya Institute of Technology, Japan), and Professor Jun Murai (Keio University, Japan). Prof. Jun Murai is known as the “Internet samurai” and, in Japan has also been called “the father of the Internet in Japan”.

We would like to thank the authors for their tremendous contributions. We would also express our sincere appreciation to the reviewers, Program Committee members and the Local Committee members for making this conference successful. Finally, we would like to express special thanks for the financial support from University of Miyazaki, Japan in making ICGEC 2015 possible.

June 2015

Thi Thi Zin
Jerry Chun-Wei Lin
Jeng-Shyang Pan
Pyke Tin
Mitsuhiro Yokota

Conference Organization

Advisory Committee Chairs

Hiromitsu Hama	Honorable Professor, Osaka City University, Japan
Pyke Tin	Former Rector, University of Computer Studies, Myanmar
Bin-Yih Liao	Kaohsiung University of Applied Sciences, Taiwan
Fujio Imai	University of Miyazaki, Japan
Jun Murai	Keio University, Japan

Advisory Committee Members

Thomas Coughlin	USA
R.M. Phatarfod	Australia
Thein Oo	Myanmar ICT Development Cooperation, Myanmar
Takashi Toriu	Osaka City University, Japan
Toshiaki Itami	University of Miyazaki, Japan
Masahiro Hamano	Asia SEED, Japan
Akinori Nishihara	Tokyo Institute of Technology, Japan

General Chairs

Kyaw Zwa Soe	Ministry of Science and Technology, Myanmar
Tatsuo Suganuma	University of Miyazaki, Japan
Vaclav Snasel	VSB-Technical University of Ostrava, Czech Republic

Program and Steering Committee Chairs

Jeng-Shyang Pan	Fujian University of Technology, China
Xin Huan Jiang	Fujian University of Technology, China
Thi Thi Zin	University of Miyazaki, Japan

Local Organization Committee Chairs

Mie Mie Mie Thet Thwin	University of Computer Studies, Yangon, Myanmar
Aye Myint	Yangon Technological University, Myanmar

Invited Session Chairs

Junzo Watada	Waseda University, Japan
Ching-Yu Yang	National Penghu University, Taiwan
Shu-Chun Chu	Flinders University, Australia

Publication Chairs

Mitsuhiro Yokota	University of Miyazaki, Japan
Jerry Chun-Wei Lin	Harbin Institute of Technology Shenzhen Graduate School, China

Finance Chairs

Mie Mie Khin	University of Computer Studies, Mandalay, Myanmar
Soe Soe Khaing	University of Technology (Yatanarpon Cyber City), Myanmar

Social Event Chairs

Myint Myint Than	Myanmar Computer Federation, Myanmar
Nay Chi Lai Lai Thein	ICTTI, Myanmar
Nang Saing Mon Khan	University of Computer Studies, Yangon, Myanmar

Program Committee Members

Noriyuki Hayashi	University of Miyazaki, Japan
Koichi Tanno	University of Miyazaki, Japan
Saw Sandar Aye	University of Information Technology, Myanmar
Moe Pwint	University of Computer Studies, Mandalay, Myanmar
Win Aye	Computer University, Mandalay, Myanmar
Thin Thu Naing	Computer University, Taunggyi, Myanmar
Aung Win	University of Technology (YCC), Myanmar
Khin Mar Lar Htun	Computer University, Hinthada, Myanmar
Myat Thidar Mon	University of Information Technology, Myanmar
Thandar Thein	Computer University, Aubin, Myanmar
Ei Ei Hlaing	Computer University, Taungoo, Myanmar
Nilar Thein	Computer University, Meikhtila, Myanmar

Myint Myint Khaing	Computer University, Pinlon, Myanmar
Nan Saw Kalayar	Computer University, Taunggyi, Myanmar
Than Nwe Aung	University of Computer Studies, Mandalay, Myanmar
Khun Oo	Myanmar Computer Federation, Myanmar
Tianhua Liu	Shenyang Normal University, China
Chang-Tsun Li	University of Warwick, UK
Sheau-Dong Lang	Central Florida University, USA
Feng-Cheng Chang	Tamkang University, Taiwan
Yuh-Yih Lu	Minghsin University of Science and Technology, Taiwan
Yueh-Hong Chen	Far East University, Taiwan
Fa-xin Yu	Zhejiang University, China
Chin-Chen Chang	Feng Chia University, Taiwan
Alexander Gelbukh	National Polytechnic Institute, Mexico
A. Fuster-Sabater	Institute of Applied Physics, Spain
Aurora Trinidad Ramirez Pozo	Federal University of Parana, Brazil
Brijesh Verma	Central Queensland University, Australia
Chien-Chang Hsu	Fu-Jen Catholic University, Taiwan
Chung-Huang Yang	National Kaohsiung Normal University, Taiwan
Chang-Shing Le	National University of Tainan, Taiwan
Chao-Chun Chen	Southern Taiwan University, Taiwan
Cheng-Hsiung Hsieh	Chaoyang University of Technology, Taiwan
Chian C. Ho	National Yunlin University of Science Technology, Taiwan
Djamila Ouelhadj	University of Portsmouth, UK
Enrique Herrera-Viedma	University of Granada, Spain
Gheorghita Ghinea	Brunel University, UK
Georgios Ch. Sirakoulis	Democritus University of Thrace, Greece
Yue Li	Nankai University, China
Huey-Ming Lee	Chinese Culture University, Taiwan
Isabel Jesus	Institute of Engineering of Porto, Portugal
Isabel L. Nunes	Universidade Nova Lisboa, Portugal
Jorge Nunez Mc Leod	Universidad Nacional de Cuyo, Argentina
Jiun-Huei Ho	Cheng Shiu University, Taiwan
Jose Alfredo F. Costa	Federal University (UFRN), Brazil
Kazumi Nakamatsu	University of Hyogo, Japan
Kwok-Yan Lam	Tsinghua University, China
Lily Lin	China University of Technology, Taiwan
Lotfi Ben Romdhane	Lotfi Ben Romdhane, Tunisia
Massimo De Gregorio	Istituto di Cibernetica, Italia
Mario Koeppen	Kyushu Institute of Technology, Japan
Michal Wozniak	Wroclaw University of Technology, Portugal
Mohsen Askari	University of Technology Sydney, Australia
Mauricio Papa	University of Tulsa, USA

Ramin Halavati	Sharif University of Technology, Iran
Shu-Chuan Chu	Cheng-Shiu University, Taiwan
Sebastian Ventura	University of Cordoba, Spain
Selva S. Rivera	Universidad Nacional de Cuyo, Argentina
Yung-Jong Shiah	Kaohsiung Medical University, Taiwan
Michael N. Vrahatis	University of Patras, Greece
Xiao-Jun Zeng	University of Manchester, UK
Xiao-Zhi Gao	Helsinki University of Technology, Finland
Yongjian Hu	South China University of Technology, China
Yu-lung Lo	Chaoyang University of Technology, Taiwan
Yusuke Nojima	Osaka Prefecture University, Japan
Zhigang Zeng	Huazhong University of Science and Technology, China
Xinpeng Zhang	Shanghai University, China
Jun-Bao Li	Harbin Institute of Technology, China
Yongping Zhang	Hisilicon Technologies Co., China
Albert B. Jeng	Jinwen University of Science and Technology, Taiwan
Wang Lidong	National Computer Network Emergency Response technical Coordination Center of China, China
Tsung-Che Chiang	National Taiwan Normal University, Taiwan
Show-Jane Yen	Mining Chuan University, Taiwan
Jung-San Lee	Feng Chia University, Taiwan
Wen-Yang Lin	National University of Kaohsiung, Taiwan
Julio Cesar Nievola	Pontificia Universidade Catolica do Parana, Brazil
Stefanos Gritzalis	University of the Aegean, Greece
Jerry Chun-Wei Lin	Harbin Institute of Technology Shenzhen Graduate School, China
Martine De Cock	Ghent University, Belgium
Ruqiang Yan	Southeast University, China
Yung-Fa Huang	Chaoyang University of Technology, Taiwan
Yau-Hwang Kuo	National Cheng kung University, Taiwan
Chia-Chen Lin	Providence University, Taiwan
Li Yao	University of Manchester, UK
Jae C. Oh	Syracuse University, USA
Estevam R. Hruschka Jr.	Federal University of Sao Carlos, Brazil
Heitor Silverio Lopes	Federal University of Technology Parana, Brazil
Shi-Jay Chen	National United University, Taiwan
Shiuh-Jeng Wang	Central Police University, Taiwan
Izzettin Temiz	Gazi University, Turkey
Akira Asano	Hiroshima University, Japan
Andri Riid	Tallinn University of Technology, Estonia
Sylvain Piechowiak	University of Valenciennes, France
Wang Feng	Kunming University of Science and Technology, China
Ming-Wen Hu	Tamkang University, Taiwan
Kun-Huang Kuo	Chienkuo Technology University, Taiwan

Yi-Nung Chung	National Changhua University of Education, Taiwan
Jenn-Kaie Lain	National Yunlin University of Science and Technology, Taiwan
Tsung-Chih Lin	Feng-Chia University, Taiwan
Chyuan-Huei Thomas Yang	Hsuan Chuang University, Taiwan
Eiji Uchino	Yamaguchi University, Japan
Shu-Hua Hua	Jinwen University of Science and Technology, Taiwan
Rung-Ching Chen	Chaoyang University of Technology, Taiwan
Shie-Jue Lee	National Sun Yat-Sen University, Taiwan
Yong Zhang	Shenzhen University, China
Luciano Sanchez	Oviedo University, Spain
Donato Impedovo	Politecnico di Bari, Italy
Hsu-Yang Kung	National Pingtung University of Science and Technology, Taiwan
Tien-Tsai Huang	Lunghwa University of Science and Technology, Taiwan
Zhiyong Zhang	Henan University of Science and Technology, China
Yuh-Chung Lin	Tajen University, Taiwan
Guan-Hsiung Liaw	I-Shou University, Taiwan
Yuchi Ming	Huazhong University of Science and Technology, China
S.N. Omkar	Indian Institute of Science, Indian
Mohammed Al Rashidi	Public Authority for Applied Education and Training (PAAET), Kuwait
Jie Jing	Zhejiang University of Technology, China
Maurice Clerc	Independent Consultant, France
K.W. Wong	City University of Hong Kong, Hong Kong
S.N. Singh	Indian Institute of Technology Kanpur, Indian
Zhihua Cui	Taiyuan University of Science and Technology, China
Marco Mussetta	Politecnico Di Torino, Italy
Ling Wang	Tsinghua University, China
Shing Chiang Tan	Multimedia University, Malaysia
Jonathan Hoyin Chan	King Mongkut's University of Technology Thonburi, Thailand
Yin Chai Wang	University Malaysia Sarawak, Malaysia
Weng Kin Lai	MIMOS Berhad, Malaysia
G. Sainarayanan	ICT Academy of Tamil Nadu, India
Sheng-Yuan Yang	St. John's University, Taiwan
Mariacarla Staffa	Universita' degli Studi di Napoli "Federico II", Italy
Wu-Chih Hu	National Penghu University of Science and Technology, Taiwan
Deng-Yuan Huang	Dayeh University, Taiwan
Pei-Yin Chen	National Cheng Kung University, Taiwan
Tsung-Han Tsai	National Central University, Taiwan

Zhijian Wu	Wuhan University, China
Sanyou Zeng	China University of Geosciences (Wuhan), China
Lei Wang	Xi'an University of Technology, China
Qingzheng Xu	Xi'an Communication Institute, China
Mu-Song Chen	Da-Yeh University, Taiwan
Hsiang-Cheh Huang	National University of Kaohsiung, Taiwan
Philippe Fournier-Viger	University of Moncton, Canada

Contents

Part I Swarm Intelligence and Its Applications

Improving the Searching Capacity of Evolved Bat Algorithm by the Periodic Signal 3
Pei-Wei Tsai, Sijing Cai, Vaci Istanda, Lyu-Chao Liao and Jeng-Shyang Pan

Image Segmentation Using KFTBES 11
Yi-Fan Li, Wei Cui, Jeng-Shyang Pan, Jun-Bao Li and Qiang Su

Evolved Bat Algorithm with Increase-Wave Strategy 21
Pei-Wei Tsai, Shunmiao Zhang, Yuan Liu, Yao He and Jeng-Shyang Pan

A New Digital Image Compression Method Using Chrestenson Transform 29
Mingyong Zhou, Zhongkan Liu and Hiromitsu Hama

Intelligent Algorithms and Their Application in Electrical Impedance Tomography 35
Xin Lu and Mingyong Zhou

Design of Multifunctional Communication Card Based on Reflective Memory 45
Shou-Da Jiang, Jia-Yu Zhao and Chang-An Wei

Computation of Large-Scale Electric Field in Free Space 55
Chang-An Wei, Qiao-Chu Cui, Lian-Lei Lin and Shou-Da Jiang

An Investigation on Maximum Entropy Estimation Based on Chrestenson Transform 65
Mingyong Zhou, Zhongkan Liu and Hiromitsu Hama

A Niching Genetic Algorithm Including an Inbreeding Mechanism for Multimodal Problems 71
Atsushi Ueno, Naoki Hagita and Tomohito Takubo

Part II Intelligent Data Security and Wearable Computing

Robust Reversible Colour Image Digital Watermarking Using Offset Coefficient Technique. 83
Ching-Yu Yang and Wen-Fong Wang

Analysis of Movement Effectiveness in Badminton Strokes with Accelerometers. 95
Wen-Fong Wang, Ching-Yu Yang and Deng-Yi Wang

A Simple and Non-invasive Measurement Method Important for Conserving Large Animals Underwater. 105
Tadamichi Morisaka, Mai Sakai, Kazunobu Kogi and Hiromitsu Hama

Characterization of Malware Detection on Android Application 113
Chit La Pyae Myo Hein and Khin Mar Myo

Enhancement of Preventing Application Layer Based on DDOS Attacks by Using Hidden Semi-Markov Model 125
Ko Ko Oo, Kyaw Zaw Ye, Hein Tun, Kyaw Zin Lin and E.M. Portnov

Heuristic Rules for Attack Detection Charged by NSL KDD Dataset . . . 137
Khaing Shwe Wutyi and Mie Mie Su Thwin

Private Cloud Deployment Model for Academic Environment Using CloudStack 155
Thinn Thu Nanig

Two Methods for Color Constancy Based on the Color Correlation Matrix 165
Takashi Toriu, Mikiya Hironaga and Naoya Hasebe

Machine Learning Techniques for Myanmar Word-Sense Disambiguation 175
Phyu Phyu Khaing and Than Nwe Aung

Part III Advance Image and Signal Processing Technology and Applications

A Graph Representative Structure for Detecting Automorphic Graphs. 189
Yu Wai Hlaing and Kyaw May Oo

A Method of Wavelength Selection of Multispectral Image for the Separation of Tumor Region 199
Kayo Yoshimoto, Hideya Takahashi, Kenji Yamada and Shigeto Yoshida

Analysis of Signal Variation Based on Path Loss in LTE Network 209
Zin Mar Myo and Myat Thida Mon

Automatic Response Analyzer with Multiple Choices in Classroom Using Image Processing and Cards 219
Hiroshi Kamada, Kazuaki Masuda and Keisuke Yamada

Design and Implementation of Real Time Vision System for Recognizing Features of Lane Boundary 229
Aung Kyaw Tun, Kyaw Zaw Ye and L.G. Gagarina

Fish Classification Based on Robust Features Selection Using Machine Learning Techniques 237
Than Thida Hnin and Khin Thidar Lynn

Multi-face Tracking with Occlusion Recovery 247
Yea-Shuan Huang and Chin-Ie Chang

Regions Labeling in Outdoor Scene Images 259
Kyawt Kyawt Htay and Nyein Aye

Segmentation of Video Image Using Changing Detection and Block Based Approach 269
Mie Mie Tin and Mie Mie Khin

Part IV IPv6 Operational Technologies and Network Applications

A Study of the IPv6 Service Availability Measurement 279
Shian-Shyong Tseng, Ai-Chin Lu, Ching-Heng Ku, Sheng-Wei Kuo and Geng-Da Tsai

Automated Home Agent Configuration for Mobile IPv6 289
Wuttipon Noopetch and Sinchai Kamolphiwong

RPKI Deployment: Risks and Alternative Solutions 299
Xiaowei Liu, Zhiwei Yan, Guanggang Geng, Xiaodong Lee, Shian-Shyong Tseng and Ching-Heng Ku

Building the Virtual Experiment Learning Activities to Facilitate IPv6 Online Training 311
Jun-Ming Su, Shian-Shyong Tseng and Ting-Wei Lin

Defense Denial-of Service Attacks on IPv6 Wireless Sensor Networks 319
Chia-Mei Chen, Sung-Chien Hsu and Gu-Hsin Lai

Filtering Duplicated Location in Tracking Traffic Data 327
Swe Swe Aung and Thinn Thu Naing

Finding Shortest Path and Transit Nodes in Public Transportation System 339
Myat Thu Zar and Myint Myint Sein

Modified Priority Search Tree for Location Aware Extracting Services 349
Linn Linn Phway and Myint Myint Sein

A Novel Research Topic Ranking System in Academic Networks 361
Thi Thi Zin, Pyke Tin and Hiromitsu Hama

SLA Guaranteed Migration Decision Model Using MCDM Approach 369
Zar Lwin Phyo

Part V Recent Progress in Computational Electromagnetic Dynamics

Evaluation of Localized Polarized Light Generated by Nano Antennas for High-Density All-Optical Magnetic Recording 383
Shinichiro Ohnuki, Keisuke Tatsuzawa, Yuta Takano, Yoshito Ashizawa and Katsuji Nakagawa

A Formulation of Electromagnetic Scattering by Defected Periodic Cylinder Array Embedded in Dielectric Slab 391
Koki Watanabe

Numerical Examination on Transmission Properties of FBG by FDTD Method 403
Angger Abdul Razak and Mitsuhiro Yokota

Numerical Examination on Validity and Usefulness for FDTD Based on Alternative Direction Implicit Method 411
Tatsuaki Mitsui and Mitsuhiro Yokota

Numerical Examination on Propagation Loss Properties in Case Human Bodies Exist Near the Transmitter Using the FDTD Method 421
Shota Sakakibara and Mitsuhiro Yokota

Measurement of Udder Surface Temperature in Cows Using Infrared Thermometer 429
Yuki Nakagawa, Noor Agha Nassary, Kiichi Fukuyama and Ikuo Kobayashi

River Flood Prediction Using Markov Model 435
Thinn Htet Htet San and Mie Mie Khin

Multiplier Changing Overclocking Method for Processor Intensive Tasks 445
Win Win Aung

Test Path Optimization Algorithm Compared with GA Based Approach 455
Aye Aye Kyaw and Myat Myat Min

Author Index 465

Part I
Swarm Intelligence and Its Applications

Improving the Searching Capacity of Evolved Bat Algorithm by the Periodic Signal

Pei-Wei Tsai, Sijing Cai, Vaci Istanda, Lyu-Chao Liao
and Jeng-Shyang Pan

Abstract In this paper, the evolved bat algorithm is improved by replacing the fixed value, which is determined by the media, with the periodic signal. The familiar periodic signal exists in the natural environment is the sine/cosine signal. We take the cosine signal in our design of improving the searching capacity of the evolved bat algorithm. Three test functions, of which the global optimum values are known, are used in the experiments. The experimental results indicate that our proposed strategy improves the searching accuracy of the evolved bat algorithm about 45.722% in average.

Keywords Evolved bat algorithm · Swarm intelligence · Periodic signal · Optimization

1 Introduction

Swarm Intelligence (SI) is a solid branch in evolutionary computing. Most algorithms in SI take tinny intelligence observed from creatures in Mother Nature as

---

P.-W. Tsai · S. Cai · J.-S. Pan(✉)
College of Information Science and Engineering,
Fujian University of Technology, Fuzhou City 350118, Fujian, China
e-mail: {pwtsai,jspan}@fjut.edu.cn, 57226503@qq.com

P.-W. Tsai · L.-C. Liao
Key Laboratory for Automotive Electronics and Electric Drive of Fujian Province,
Fujian University of Technology, Fuzhou City 350118, Fujian, China
e-mail: achao@fjut.edu.cn

V. Istanda
Council of Indigenous Peoples, Executive Yuan, New Taipei City 24220, Taipei, Taiwan
e-mail: biungsu@yahoo.com.tw

J.-S. Pan
Innovative Information Industry Research Center, Shenzhen Graduate School,
Harbin Institute of Technology, Shenzhen 518055, China

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_1

the guidance for dragging the artificial agents to find new solutions in optimization. Many algorithms in SI are proposed one after another in recent years. For instant, Cat Swarm Optimization (CSO) [1], [2] utilizes the seeking mode to simulate the cat is observing the environment and the tracing mode to imitate the coursing behavior. In addition, many improved version of CSO has been presented in recent years and its applications are also proposed one after another. For example, Temel et al. utilize CSO with wavelet transform to maximize the sensing coverage in Wireless Sensor Network (WSN) [3], Pappula and Ghosh employ CSO in linear antenna array synthesis [4], and Yang et al. use CSO with the limited memory Broyden–Fletcher–Goldfarb–Shanno with boundaries (L-BFGS-B) method to calibrate the medical image registration [5]. On the other hand, Evolved Bat Algorithm (EBA) [6] is proposed in 2012 for solving numerical optimization problems by simulating the behavior of bats searching for prey in the space. EBA is an inspired algorithm from the original Bat Algorithm (BA) [7]. The related applications of BA include structural optimization [8]; Niknam et al. utilize BA in finding the linear supply function equilibrium of GENCOs in the competitive electricity market [9]; and BA is also used in solving multiobjective optimization problems in reserve constrained dynamic environmental economic dispatch [10].

In this paper, we focus on further improving the searching capacity of EBA by leading the periodical signal in the searching operation. The periodical signal brings in the vibration to the artificial agents and provides larger diversity of the individuals. To test the influence of leading the periodic signal in the searching process, three test functions are employed in the experiments. The experimental results indicate that our proposed strategy improves the searching accuracy of the evolved bat algorithm about 45.722% in average. The rest of the paper is constructed as follows: a brief review of EBA is given in section 2, our proposed strategy is described in detail in section 3, the experiments and the experimental results are given in section 4, and the conclusions are made at last.

2 Literature Review

In 2012, Tsai et al. present EBA by redesigning the movement operation of the original BA and simplifying the control parameters required by the algorithm. The distance calculation based on the active sonar system is taken into account in the artificial agents' movement. The key element of the movement of the artificial agent in EBA is processed by Eqs. (1)-(2):

$$D = 0.17 \cdot \Delta T \tag{1}$$

$$x_i^t = x_i^{t-1} + D \tag{2}$$

where D denotes the distance between the current position to the prey, ΔT is a random variable and $\Delta T \in [-1, 1]$, x_i represents a solution obtained by the

i^{th} artificial agent, and t stands for the count of the iterations. In addition, every artificial agent has half chance to be further processed by the random walk operation by Eq. (3):

$$x_i^{t_R} = \beta \cdot (x_{best} - x_i^t) \tag{3}$$

where $x_i^{t_R}$ indicates the artificial agent after the random walk operation, β is a random number in the range of [0, 1], and x_{best} stands for the near best solution obtained in the past iterations.

The process of EBA is briefly reviewed as follows:

- Step 1. Initialization: Spread the artificial agents into the solution space.
- Step 2. Movement: Change the solutions of the artificial agent by Eqs. (1)-(2). Generate a random number to decide whether takes the random walk process. If the random walk process is taken into the operation, Eq. (3) is employed.
- Step 3. Evaluation: Evaluate the fitness values of the artificial agents and update the near best solution found in the past iterations.
- Step 4. Termination: If the termination condition is satisfied, terminate the program and output the near best solution; otherwise, go back to step 2 and repeat the process.

3 Our Proposed Strategy

When presenting EAB, Tsai et al. choose the air to be the medium for carrying the sound wave. Based on the known truth, the sound speed spreads in the air is 340 meters per second. In addition, the sound wave actually travels twice far from the wave source to the prey and the traveling direction is taken in the consideration. Hence, the travel distance in EBA is a fixed range in [-0.17, 0.17]. Although the users may set the range in different values by their own design, the diversity pumped into the artificial agents is still not strong and powerful for exploring the solution space. To overcome this drawback, we employ the periodic signal to replace the fixed limit. The most common periodic signal, which can be detected in the nature environment is the cosine signal. Hence, we inject the cosine signal into the movement of the algorithm to provide larger diversities. In our proposed strategy, the movement of the artificial agent is processed by Eqs. (4)-(5):

$$Wave = A \cdot \cos(\omega t) \tag{4}$$

$$x_i^t = x_i^{t-1} + Wave \cdot \Delta T \tag{5}$$

where $Wave$ denotes the periodical signal employed in the operation, and A is the amplitude of the signal.

In our design, a full cycle of the injected periodical signal is defined by Eq. (6):

$$T = \frac{1}{t_i} \tag{6}$$

where T stands for the cycle and t_i indicates a set of number of iterations.

An example of the injected periodical signal is shown in Fig. 1.

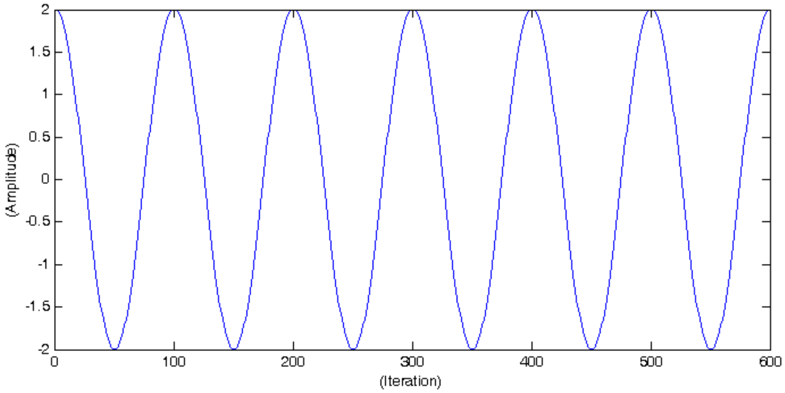


Fig. 1 Sample of the injected periodical signal for our proposed strategy.

As shown in Fig. 1, our proposed strategy utilize the cosine signal to increase the diversity of the searching range. The example given here uses $A = 2$ as the amplitude, and the length of a full cycle of the signal crosses 100 iterations.

4 Experiments and Experimental Results

To test the impact to the searching capacity brought by our proposed strategy, three test functions with known global optimum are used in the experiments. The test functions are listed in Eqs. (7)-(9):

$$f_1(X) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \tag{7}$$

$$f_2(X) = 20 + e^1 - 20 \cdot e^{-0.2\sqrt{\sum_{i=1}^{n}\frac{x_i^2}{n}}} - e^{\sum_{i=1}^{n}\frac{\cos(2\pi \cdot x_i)}{n}} \tag{8}$$

$$f_3(X) = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 10\cos(2\pi \cdot x_i) + 10 \cdot n \tag{9}$$

As the same as most engineering applications, the goal of the optimizations of all test functions is to find the minimum value of the benchmark. The dimension of all test functions is set to 30, all experiments are repeated for 25 runs for finding the average results, the total iteration number is set to 6000 for Eq. (7) and 5000 for Eqs. (8)-(9), and the initial range of these functions are set to [-600, 600], [3, 13], and [-5.12, 5.12]. The parameter setting of the original EBA and EBA with our proposed strategy is given in Table 1.

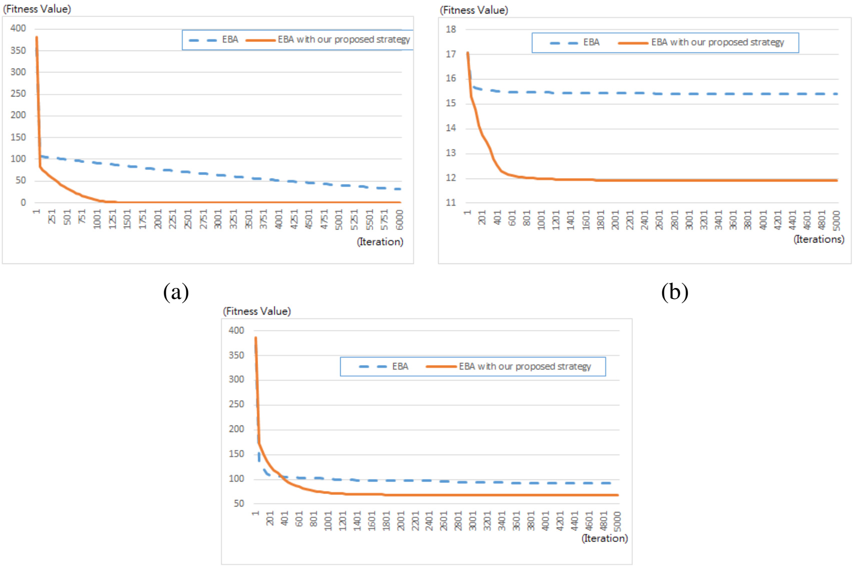
Table 1 Parameter setting for EBA and EBA with our proposed strategy.

	EBA [6]	EBA with our proposed strategy
Fixed *r*	0.5	0.5
ΔT	[-1, 1]	[-1, 1]
A	NA	2

The final fitness values of all test functions are listed in Table 2.

Table 2 The average results over 25 runs.

	EBA [6]	EBA with our proposed strategy
$f_1(X)$	3.3647×10^{0}	2.5255×10^{-2}
$f_2(X)$	1.5395×10^{1}	1.1911×10^{1}
$f_3(X)$	7.9495×10^{1}	6.7390×10^{1}



(a) (b)

(c)

Fig. 2 The fitness values of all test functions: (a) for $f_1(X)$, (b) for $f_2(X)$, (c) for $f_3(X)$.

Fig. 2 shows the convergence of the fitness values of all test functions. It is obvious that the convergence of EBA with our proposed strategy presents faster convergence and higher accuracy in finding the near best solutions.

5 Conclusions

In this paper, we propose a strategy, which injects a periodical signal into the process of EBA. We use the cosign wave, which is the most common periodical signal can be detected in the nature environment, to reformat the equations employed in the movement process in EBA. It results in the increasing of the diversity of artificial agents when searching for new solutions. To test the usability of this strategy, three fitness functions with known global optimal are used in the experiments. The accuracy of finding the near best solution of our proposed strategy is compared with the conventional EBA. The experimental results indicates that our proposed strategy significantly improves the searching accuracy about 45.722% in average.

In the future work, we will focus on improving the searching efficiency of the swarm intelligence algorithm. Furthermore, some industrial applications such as combining EBA algorithm with the intelligent information hiding algorithms [11], [12], [13], [14], [15] may be taken into considered for combining with the swarm intelligence algorithms.

Acknowledgement The authors would like to thank Fujian University of Technology in China, for the financial support of this research.

References

1. Chu, S.-C., Tsai, P.-W.: Computational Intelligence Based on the Behavior of Cats. International Journal of Innovative Computing, Information and Control **3**(1), 163–173 (2007)
2. Chu, S.-C., Tsai, P.-w., Pan, J.-S.: Cat swarm optimization. In: Yang, Q., Webb, G. (eds.) PRICAI 2006. LNCS (LNAI), vol. 4099, pp. 854–858. Springer, Heidelberg (2006)
3. Temel, S., Unaldi, N., Kaynak, O.: On Deployment of Wireless Sensors on 3-D Terrains to Maximize Sensing Coverage by Utilizing Cat Swarm Optimization with Wavelet Transform. IEEE Transactions on Systems, Man, and Cybernetics: Systems **44**(1), 111–120 (2014)
4. Pappula, L., Ghosh, D.: Linear antenna array synthesis using cat swarm optimization. International Journal of Electronics and Communications (AEÜ) **68**, 540–549 (2014)
5. Yang, F., Ding, M., Zhang, X., Hou, W., Zhong, C.: Non-rigid multi-modal medical image registration by combining L-BFGS-B with cat swarm optimization. Information Sciences, in press (2015)
6. Tsai, P.-W., Pan, J.-S., Liao, B.-Y., Tsai, M.-J., Vaci, I.: Bat Algorithm Inspired Algorithm for Solving Numerical Optimization Problems. Applied Mechanics and Materials **148–149**, 134–137 (2012)

7. Yang, X.-S.: A new metaheuristic bat-inspired algorithm. In: González, J.R., Pelta, D.A., Cruz, C., Terrazas, G., Krasnogor, N. (eds.) NICSO 2010. SCI, vol. 284, pp. 65–74. Springer, Heidelberg (2010)
8. Hasançebi, O., Teke, T., Pekcan, O.: A bat-inspired algorithm for structural optimization. Computers and Structures **128**, 77–90 (2013)
9. Niknam, T., Sharifinia, S., Azizipanah-Abarghooee, R.: A new enhanced bat-inspired algorithm for finding linear supply function equilibrium of GENCOs in the competitive electricity market. Energy Conversion and Management **76**, 1015–1028 (2013)
10. Niknam, T., Azizipanah-Abarghooee, R., Zare, M., Bahmani-Firouzi, B.: Reserve Constrained Dynamic Environmental/Economic Dispatch: A New Multiobjective Self-Adaptive Learning Bat Algorithm. IEEE Systems Journal **7**(4), 763–776 (2013)
11. Zhao, D.-N., Xie, W.-X., Lu, Z.-M.: High Efficiency Reversible Data Hiding for Two-stage Vector Quantization Compressed Images. Journal of Information Hiding and Multimedia Signal Processing **5**(4), 625–641 (2014)
12. Ngo, N.M., Unoki, M., Miyauchi, R., Suzuki, Y.: Data Hiding Scheme for Amplitude Modulation Radio Broadcasting Systems. Journal of Information Hiding and Multimedia Signal Processing **5**(3), 324–341 (2014)
13. Li, P., Kong, Q., Ma, Y.: Image Secret Sharing and Hiding with Authentication Based on PSNR Estimation. Journal of Information Hiding and Multimedia Signal Processing **5**(3), 353–366 (2014)
14. Kuo, W.-C., Chang, S.-Y.: Hybrid GEMD Data Hiding. Journal of Information Hiding and Multimedia Signal Processing **5**(3), 420–430 (2014)
15. Marin, J., Shih, F.Y.: Reversible Data Hiding Techniques Using Multiple Scanning Difference Value Histogram Modification. Journal of Information Hiding and Multimedia Signal Processing **5**(3), 451–460 (2014)

Image Segmentation Using KFTBES

Yi-Fan Li, Wei Cui, Jeng-Shyang Pan, Jun-Bao Li and Qiang Su

Abstract We present a supervised algorithm to improve the image segmentation algorithm based on texture and boundary encoding. Our method is due to the analysis of the implementation and result of the TBES algorithms, and we increase the adaptability of the TBES algorithms. Through constructing the train dataset with fine-class segmentation, our method adaptively distribute the optimum segmentation standard to each image using Kernel Fisher algorithms. We also compare our method to other similar popular algorithms and our method achieves the state-of-the-art results on Berkeley Segmentation Dataset.

Keywords Image segmentation · TBES algorithm · Kernel Fisher

1 Introduction

Image segmentation is always a long-term and typical problem in computer vision, and its main difficulties are boundary location of regions and the degree of regional integration. The models and methods of image segmentation learn from the process of eye recognition, so many algorithms increase the accuracy of image

---

Y.-F. Li · W. Cui · J.-S. Pan(✉) · Q. Su
Innovative Information Industry Research Center, Shenzhen Graduate School,
Harbin Institute of Technology, Shenzhen 518005, China
e-mail: {lyf3f,hitsuqiang}@163.com, tianshiwu1989@126.com,
jengshyangpan@gmail.com

J.-S. Pan
College of Information Science and Engineering,
Fujian University of Technology, Fuzhou 350118, China

J.-B. Li
Department of Automatic Test and Control,
Harbin Institute of Technology, Harbin, China
e-mail: lijunbao@hit.edu.cn

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_2

segmentation by imitate the eye recognition. Therefore image segmentation has the following problems and challenge.

The first challenge is the diversity of scenes and objects. Image segmentation is to divide the image into several parts with inherent characteristic, and the diversity of scenes and objects greatly increase the complexity and difficulty of creating model. The quality of model resolves the result of algorithm, while models change along the variety of scenes.

The second challenge is the diversity of evaluation criteria. There are a lot of segmentation methods on the same image by people, and especially the image contains wealth of contents. Meanwhile the same person may give different results along the change of time and knowledge. Therefore the lack of unified parameter model results in the insufficiency of methods to ameliorate the outcome.

According to the above-mentioned problem, algorithms are manifold on image segmentation. However, the vast majority of methods firstly detect the possible boundary, and then graphic division based the edge feature. Customarily edge detection is grounded on the brightness feature, such as Canny operators, Sobel operators [2]. They use the low-level feature to detect the location of the edge, while others use mid-level and high-level feature to locate the boundary, such as Level Set [3]. But the high-level feature is pertinent to the special dataset.

Here we detect the image boundaries using local brightness, color and texture clues [4], and hence we can detect the strong and weak edges to the utmost. Based on the clues of boundary, Superpixels algorithm [5] cut the image into numbers of regions with identical inherent characteristics named Superpixel. Thus we can view regions as special pixels to decrease the complexity. Among the region cluster algorithms, image segmentation with adaptive texture and boundary encoding algorithm (TBES) [6] achieve the optimum, so it is used in this thesis to merge regions. What's more, the parameters need to be adjusted in this method are few, and the final result is decide by only one parameter named distortion. Hence it is easy to adjust the segmentation result and achieve the target level. In order to adaptively determine the distortion for every image, we use Kernel Fisher algorithm [18], [19], [20] to classify each image. Firstly we construct the training dataset in which images are endued the optimum distortion by human, and then allocate the appropriate distortion to the images in the testing dataset based on the training dataset. For an input image, we firstly divide it into regions using Superpixels algorithm, and then merge regions based on the result of classification using Kernel Fisher algorithm. In this way, image segmentation not only achieves good result, but also provides materials for other image process, such as scene labeling [7]. We also compare our method with other popular segmentation algorithm.

This paper is organized as follows: Section 2 briefly reviews the recently popular and relevant image segmentation algorithm. Section 3 introduces the main process and result of Superpixels algorithm. Section 4 depicts details and effect of TBES algorithm. Section 5 describes our method how to improve TBES algorithm by using Kernel Fisher algorithm. Section 6 gives the experiment results on Berkeley Segmentation Dataset and compares to other existing algorithms. Section 7 concludes the paper.

2 Related Work

In the recent years, there are some important and effective methods and models on image segmentation. Normalized Cut (NC) [8] and F&H [9] view segmentation as graph-cut problem, and Mean Shift (MS) [10] focuses on the density distribution with estimated empirical evaluation. Boundary shape feature convey import information, so segmentation algorithms based boundary shape [11], [12]. Since local clues have its boundedness for global analysis, some methods utilize various level of characteristic [13], [14]. However, there is hardly a balance between subjective standard and objective standard to evaluate the segmentation and promote the improvement on methods. Some methods are inclined to the objective standard such as lossy minimum description length (MDL). These methods pursue the optimum result through encoding the texture and boundary of region, and calculating the minimum encoding length, such as compression-based texture merging (CTM) [15]. But CTM algorithm exploits overlapped windows to extract the texture and use Huffman code to encode the texture and boundary. The segmentation can not reflect the global entropy and ignore the spatial connectivity and boundary smoothness.

3 Superpixels Algorithm

This section describes the initial segmentation using Superpixels algorithm in Section 3.1 and the result in Section 3.2.

3.1 Boundary Detection

Given an image ***I***, we firstly detect the boundary using local brightness, color and texture feature [16]. Through setting a circular mask whose radius is *r*, we extract the features of the center pixel *p(x, y)* in eight orientations. The approach to use the feature is the calculation of the gradient difference between two semi-circle produced by cutting the circle mask. As shown in Fig. 3.1 (a), we could calculate the gradient change in the orientation which is vertical to the diameter, and as a matter of fact we look at each pixel for local discontinuities over a range of orientations and scales. Brightness and color features are extracted under the LAB model, while the texture feature is extracted under the gray model. The formulas of gradient computation are same for three features:

$$1.\ x^2(g,h) = \sum (g_i - h_i)^2 \tag{1}$$

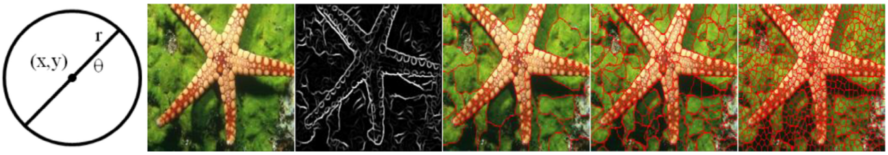
In formula (1), g_i and h_i are the kernel density estimates of the distributions of pixel luminance and chrominance using histograms [1] in each disc half. For the brightness and color features, the number of bins is 32, while the number is 64 for texture feature. Then we calculate the probability for each pixel as the element of boundary.

$$probability = 1 - e^{-\frac{x^2}{2\sigma^2}} \tag{2}$$

In formula (2), x^2 is computed by formula (1) and is the sum of three features. As shown in the Fig. 1 (b), through the calculation, we can get the probability of the existing boundary named *pb (x, y, θ)* for each pixel in the orientation θ. If there exists boundary, the gradients of features will change fiercely with the representation of high probability. Compared to the Canny algorithm [2] and 2MM algorithm [17], Superpixels algorithm detect the weak boundary and strong boundary better.

3.2 Segmentation Result

We construct the initial regions using the Superpixels algorithm and we can set the number of initial regions on the basis of our targets. From the Fig. 1 (d) ~ (f), it is found that the loss of boundary increases with the decrease of the number of initial regions (*K* is the number of segmentation regions). Through the experiments on the images, it is found that the strong boundary is the framework and the weak boundary is the detail. Superpixels algorithm could well distinguish the strong boundary and the weak boundary, so that it prepares the ground for good segmentation with alternative number.



(a) Mask (b) Original (c) Boundary (d) *K*=40 (e) *K*=200 (f) *K*=600

Fig. 1 The mask and the results of Superpixels algorithm

4 Region Cluster Algorithm

Region cluster algorithm based on minimum description length (MDL) merges neighbor regions though pursuing the least encoding length with encoding the texture and boundary of regions. Here we adopt the adaptive texture and boundary encoding algorithm (TBES) [6].

4.1 TBES Algorithm

The initial task of cluster algorithm based on the MDL is the manner how to encoding the features of regions, and the manner is supposed to be closely related to the basic features of regions, such as the shape, color, boundary and relation between regions.

The exacted features of TBES algorithm mainly are color and boundary. Using the non-overlapping interior square windows over a range of size, we extract the color feature column vectors under the LAB mode. As shown in Fig. 2 (a), we extract the color features of the target region R using non-lapping interior mask whose elements belong to the same region, and the scale of the mask is variable to avoid the lack of the feature vectors for some small regions. Then the feature vectors are improved using the PCA algorithm to reduce the dimension, and represented by mean and covariance. The expected number of bit needed to code the texture feature of a region is given by:

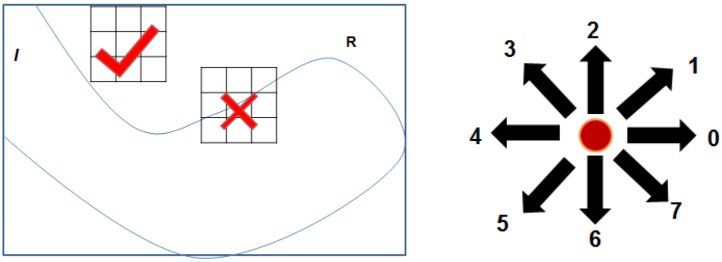
$$L_\varepsilon(\hat{X}) = \frac{D}{2} log_2\, det\, (I + \frac{D}{\varepsilon^2}\Sigma) + \frac{N}{2} log_2\, det\, (I + \frac{D}{\varepsilon^2}\Sigma) + \frac{D}{2} log_2\, det\, (I + \frac{||\mu||^2}{\varepsilon^2}) \quad (3)$$

In formula (3), D is the dimension of the feature vector, and here we set it as 8. Σ is the covariance of all the feature vectors in a region while μ is the mean vector of all the feature vectors in a region. N is the number of the pixels of the region and ε is the lossy evaluation which determines the degree of the merging.

For the boundary of regions, encoding method [21] is using the starting point and the direction in turn. As shown in Fig. 2 (b), the eight numbers represent the eight directions, and turn the boundary into the digit sequence with the start point. However, the used digit sequence is the difference sequence since this sequence produces a number of zero elements to reduce the complexity. The expected number of bit needed to obtain the boundary feature of a region is given by:

$$B(R) = -\sum_{i=0}^{7} \#(\Delta o_t = i)\, log_2(P[\Delta o = i]) \quad (4)$$

In formula (4), $\#(\Delta o_t = i)$ is the storage capacities of difference for each direction, while $P[\Delta o = i]$ is the number of the boundary i.



(a) (b)

Fig. 2 The window to extract feature and the boundary encoding

To determine whether merge the neighbor regions R_i and R_j, we use the following merging standard:

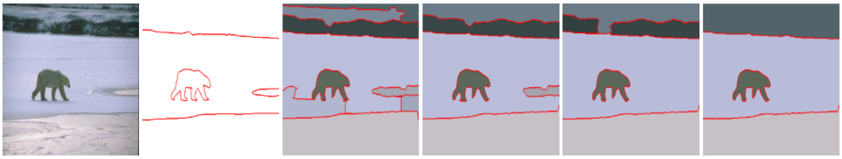
$$\Delta \mathrm{L}_{w,\varepsilon}(R_i, R_j) = L_{w,\varepsilon}^{S}(\mathcal{R}) - L_{w,\varepsilon}^{S}\left((\mathcal{R}\backslash\{R_i, R_j\})\cup\{R_i \cup R_j\}\right)$$

$$= L_{w,\varepsilon}(R_i) + L_{w,\varepsilon}(R_j) - L_{w,\varepsilon}(R_i \cup R_j) + \frac{1}{2}\left(B(R_i) + B(R_j) - B(R_i \cup R_i)\right) \qquad (5)$$

In the formula (5), R_i, R_j represent two neighbor regions, and $\mathcal{R}$ is the merging region combines of region R_i and R_j. if $\Delta L_{w,\varepsilon}(R_i, R_j) > 0$, then we could merge the two regions. Otherwise, keep intact.

4.2 Cluster Result

Through the variety of the lossy evaluation ε, the extent of segmentation changes gradually. Fig. 3 (c) ~ (f) show the different extent of segmentation, and the degree of cluster amplifies with the increase of ε. Therefore we could prescribe ε based on the target segmentation to provide convenience to further applications.



(a) Original (b) Human (c) ε=25 (d) ε=50 (e) ε=100 (f) ε=175

Fig. 3 The result of TBES algorithm with different *ε*

5 Kernel Fisher Improvement

The lossy evaluation ε adjusts the degree of the cluster, and could be determined adaptively using the Kernel Fisher algorithm. In this section, we introduce the improvement of TBES algorithm using the Kernel Fisher algorithm

Among the classification algorithms, Kernel Fisher algorithm gains good classification performance, and achieves high accuracy in the human face recognition. Here we use Kernel Fisher to determine the lossy evaluation ε for the test image. Firstly, we construct the target train dataset through setting the optimal lossy evaluation ε for each images from the train dataset with the evaluation standards. Thus we can gain k class data $X_i = \{x_1^i, x_2^i, x_3^i, x_4^i, x_5^i, \cdots\cdots\}$, $i \in [1, k]$. Using the Kernel Method to establish a new mapping $\emptyset$ from the original space to the kernel space, the mapping function is given by:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d, d = 1, 2, \cdots\cdots, n \qquad (6)$$

Through the transformation, $\mathrm{x_i}$ is mapped into the kernel space as $\mathrm{KX}_i = \{\emptyset(x_1^i), \emptyset(x_2^i), \emptyset(x_3^i), \cdots\cdots\}$, $\mathrm{i} \in [1, \mathrm{k}]$. According to the Fisher algorithm, we need to calculate the project direction ω, the within class variance S_W, and the between class variance S_B. Since the current space is the kernel space, S_W and

S_B are supposed to be computed using the vectors after mapping. The projection direction $\omega^{\emptyset}$ in the kernel space is obtained by the following formula:

$$W_{OPT}^{\emptyset} = argmax_{W^{\emptyset}} \frac{\left|(W^{\emptyset})^T S_B^{\emptyset} W^{\emptyset}\right|}{|(W^{\emptyset})^T S_W^{\emptyset} W^{\emptyset}}$$

$$= argmax_{W^{\emptyset}} \frac{|\alpha KZK\alpha|}{|\alpha KK\alpha|}$$

$$= [w_1^{\emptyset}, w_2^{\emptyset}, w_3^{\emptyset}, w_4^{\emptyset} \ldots\ldots w_k^{\emptyset}] \tag{7}$$

In formula (7), $w_i^{\emptyset}$ is the best projection for the class i. For the images in the test dataset, we firstly transform them to kernel space, and then map them to each class using the corresponding $w_i^{\emptyset}$. The final label for each image is determined by formula (8).

$$y^{\emptyset} = \overset{argmax}{i} (w_i^{\emptyset} x^{\emptyset} + b_i) \tag{8}$$

In the formula (8), b_i is determined by the $S_W^{\emptyset}$ and the $S_B^{\emptyset}$ of class i, and the final label is the label corresponding to the maximum of $y^{\emptyset}$ over all the class. Using the feature of gray model of images, we gain the good segmentation with the adjustment of parameters and the evaluation of the result is shown in the Section 6.

6 Experiment Results

This section introduces our experiment on the Berkeley Segmentation Dataset (BSD) which is presented by David Martin and Charless Fowlkes of the image segmentation research group of UC Berkeley. In order to evaluate the result well compared with the similar algorithms, we use the Probabilistic Rand Index (PRI) and the Variation of Information (VOI) to estimate the results.

BSD 500 contains the training dataset and testing dataset and each dataset contains 200 images. We firstly utilize the 200 images in the file named test, and choose the optimum lossy evaluation ε. According to the result of many experiments, we set 8 classes between whom the segmentations change obviously. The corresponding evaluations ε are 25, 50, 75,100, 125, 150, 175 and 200. Then we use the features of gray mode of images, since it is found that the image segmentation is close to the structure of the object distribution in the image. Using the Kernel Fisher algorithm, we distribute the images in the file named train with the corresponding lossy evaluation ε.

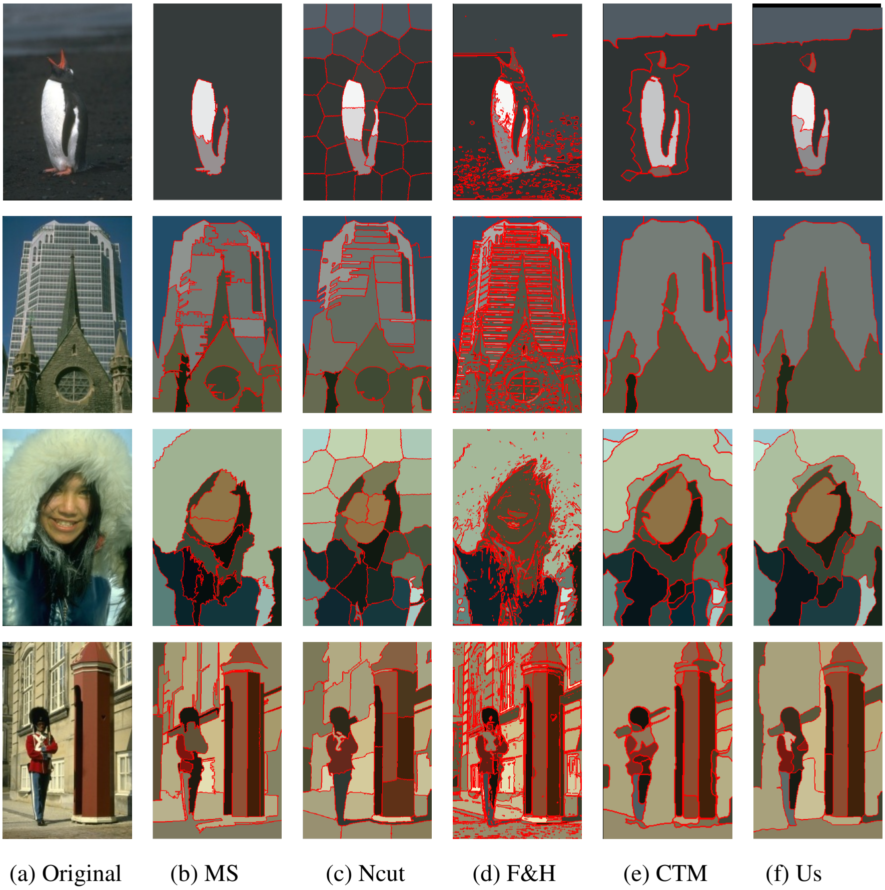

(a) Original (b) MS (c) Ncut (d) F&H (e) CTM (f) Us

Fig. 4 The segmentation of various algorithms

Table 1 Comparision of PRI and VOI for various algorithms on BSD

Index/Method	MS	NCUT	F&H	CTM	TBES	KFTBES	Human
PRI (higher is better)	0.7914	0.7139	0.7792	0.7624	0.8000	0.8047	0.8867
VOI (lower is better)	1.9146	2.6693	2.4404	2.3005	1.7004	1.7003	1.0071

We compare our performance with other image segmentation algorithm, such as “MS”, “NCUT”, “F&H”, “CTM”, “TBES” and human segmentation. The performances of other algorithms are achieved by selecting the optimum result over various parameters. From Table 1 we could find that our method achieves the best performance compared to other popular segmentation methods. Fig. 4 shows the image segmentations of different algorithms of several images, and our method adaptively chooses the degree of segmentation instead of frequent adjustment of parameters like other algorithms.

7 Conclusion

We have proposed an improved method for TBES algorithm to achieve adaptive setting of the lossy evaluation. The algorithm used Kernel Fisher algorithm to classify the images with the corresponding lossy evaluation. Considering the result is determined by the training dataset, we could adjust the parameters of the training dataset to achieve the target result, and use the result to other image process such as scene labeling. However, the number of dataset in our experiment is not comprehensive, so we need to do more experiments to further affirm the effect of our algorithm.

References

1. Krinidis, S., Krinidis, M., Chatzis, V.: An unsupervised image clustering method based on EEMD image histogram. Journal of Information Hiding and Multimedia Signal Processing **3**(2), 152–163 (2012)
2. Canny, J.F.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **6**, 679–698 (1986)
3. Osher, S., Sethian, J.A.: Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. Journal of computational physics **79**(1), 12–49 (1988)
4. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(5), 530–549 (2004)
5. Ren, X., Malik, J.: Learning a classification model for segmentation. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, pp. 10–17. IEEE (2003)
6. Rao, S.R., Mobahi, H., Yang, A.Y., Sastry, S., Ma, Y.: Natural image segmentation with adaptive texture and boundary encoding. In: Zha, H., Taniguchi, R.-i., Maybank, S. (eds.) ACCV 2009, Part I. LNCS, vol. 5994, pp. 135–146. Springer, Heidelberg (2010)
7. Farabet, C., Couprie, C., Najman, L., et al.: Scene parsing with multiscale feature learning, purity trees, and optimal covers. arXiv preprint arXiv:1202.2160 (2012)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(8), 888–905 (2000)
9. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision **59**(2), 167–181 (2004)
10. Comaneci D., Shift P. M.M.: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(5) (2002)
11. Arbelaez, P.: Boundary extraction in natural images using ultrametric contour maps. In: Conference on Computer Vision and Pattern Recognition Workshop. CVPRW 2006. IEEE, pp. 182–182 (2006)
12. Ren, X., Fowlkes, C.C., Malik, J.: Learning probabilistic models for contour completion in natural images. International Journal of Computer Vision **77**(1–3), 47–63 (2008)

13. Yu S.X.: Segmentation induced by scale invariance. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 444–451. IEEE (2005)
14. Ren, X., Fowlkes, C.C., Malik, J.: Scale-invariant contour completion using conditional random fields. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 2, pp. 1214–1221. IEEE (2005)
15. Yang, A.Y., Wright, J., Ma, Y., et al.: Unsupervised segmentation of natural images via lossy data compression. Computer Vision and Image Understanding **110**(2), 212–225 (2008)
16. Belongie, S., Carson, C., Greenspan, H., et al: Color-and texture-based image segmentation using EM and its application to content-based image retrieval. In: Sixth International Conference on Computer Vision, pp. 675–682. IEEE (1998)
17. Pan, J.S., Li, J.B., Lu, Z.M.: Adaptive quasiconformal kernel discriminant analysis. Neurocomputing **71**(13), 2754–2760 (2008)
18. Li, J.B., Pan, J.S., Lu, Z.M.: Kernel optimization-based discriminant analysis for face recognition. Neural Computing and Applications **18**(6), 603–612 (2009)
19. Li, J.B., Pan, J.S., Lu, Z.M.: Face recognition using Gabor-based complete Kernel Fisher Discriminant analysis with fractional power polynomial models. Neural Computing and Applications **18**(6), 613–621 (2009)
20. Chen, H., Liu, B.B., Luo, H., et al.: Fast image artistic style learning using twin-codebook vector quantization. Journal of Information Hiding Multimedia Signal Process **3**(1), 66–70 (2012)

Evolved Bat Algorithm with Increase-Wave Strategy

Pei-Wei Tsai, Shunmiao Zhang, Yuan Liu, Yao He and Jeng-Shyang Pan

Abstract In this paper, a mixture signal, which is composed of a periodical signal and an increasing level Direct Current (DC) signal, is used in this paper to amplify the diversity brought to the artificial agents in Evolved Bat Algorithm (EBA). In order to test the accuracy on finding the near best solutions, three test functions with known global optimum are used in the experiments. The experimental results indicate that our method improves the searching accuracy of EBA about 47.74 percent in average.

Keywords Evolved bat algorithm · Swarm intelligence · Optimization

1 Introduction

Evolved Bat Algorithm (EBA) [1] is one of the newly released algorithm for solving numerical optimization problems in Swarm Intelligence (SI) in recent years. Like other algorithms in SI, EBA takes the tinny intelligence exists in bat's natural instinct and uses its behavior of hunting for the prey to form the mathematical model. The experimental results indicate that EBA presents better accuracy and performance than the original Bat Algorithm (BA) [2]. However, EBA uses a fixed constant in the process of presenting the new solutions. The fixed constant sometimes forms a bottleneck for the artificial agents to discover better solutions. In order to overcome this drawback, we use a mixture signal in the searching

---

P.-W. Tsai · S. Zhang · Y. Liu · Y. He · J.-S. Pan(✉)
College of Information Science and Engineering,
Fujian University of Technology, Fuzhou City 350118, Fujian, China
e-mail: {pwtsai,liuyuan,jspan}@fjut.edu.cn, {zshunmiao,wonderkewen}@163.com

J.-S. Pan
Innovative Information Industry Research Center,
Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, China

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_3

process to inject higher diversity to the artificial agents. Although the conventional EBA has the drawback mentioned above, it is still successfully used in finding feasible solutions for the LMI fuzzy Lyapunov nonlinear time-delay system controller [3]. And BA is used in structural optimization [4], and multiobjective optimization [5].

Besides EBA, many algorithms in SI are also proposed one after another, and their applications include many fields in engineering. For example, Interactive Artificial Bee Colony (IABC) [6] is successfully used in the passive continuous authentication system [7]; Cat Swarm Optimization (CSO) [8], [9], is modified and applied in the scheduling problem [10], deploying Wireless Sensor Networks (WSN) [11], [12], designing linear antenna array synthesis [13], and constructing the medical image registration [14].

After revisiting the way EBA produces new solutions, we replace the fixed constant, which represents the medium, by a mixture signal composed of a periodical signal and a level-increasing dc signal. The vibration is brought into the process to the artificial agents, and the diversity is also increased. In order to test the influence caused by our change, we use three test functions, of which the global optimum values are known, in the experiment. The experimental results indicate that our proposed method improves the searching accuracy of EBA about 47.74 percent in average. The rest of this paper is constructed as follows: in section 2, a brief review on EBA is given, the detail of our proposed method is revealed in section 3, the experiments and the experimental results are given in section 4, and the conclusions are made in the last section.

2 Literature Review

Tsai et al. present the conventional EBA in 2012 by redesigning the operation of which BA finds new solutions. The control parameters are much simplified in EBA than in BA. The distance calculation in the active sonar system is employed to be part of the operation for moving the artificial agents. The movement of the artificial agents in EBA can be depicted by Eqs. (1)-(2):

$$D = 0.17 \cdot \Delta T \tag{1}$$

$$x_i^t = x_i^{t-1} + D \tag{2}$$

where $\Delta T \in [-1, 1]$ is a random variable, D denotes the distance between the current position to the prey, x_i stands for a solution obtained by the i^{th} artificial agent, and t identifies the number of current iteration.

Besides the regular movement, EBA also provides 50% chance for the artificial agents to further take the random walk process. The random walk process is operated by Eq. (3):

$$x_i^{t_R} = \beta \cdot (x_{best} - x_i^t) \tag{3}$$

where $x_i^{t_R}$ represents the artificial agent after the random walk operation, β denotes a random number in the range of [0, 1], and x_{best} is the near best solution obtained in the past iterations.

The diagram of the conventional EBA is depicted in Fig. 1:

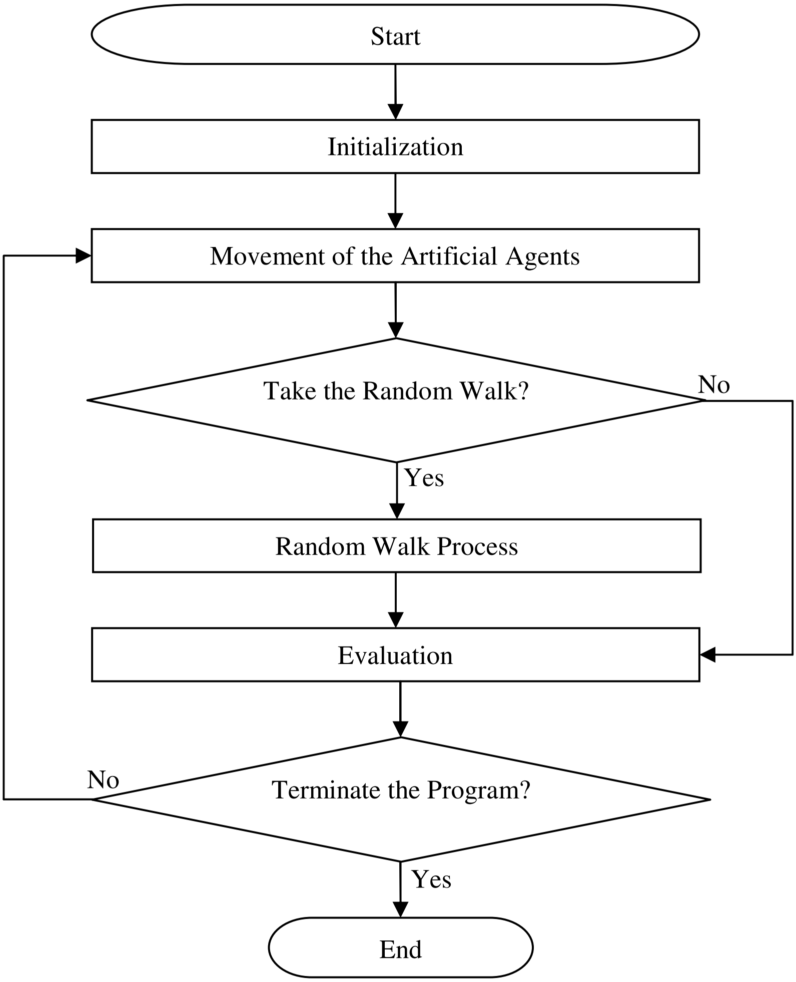

Start
Initialization
Movement of the Artificial Agents
Take the Random Walk?
No
Yes
Random Walk Process
Evaluation
No
Terminate the Program?
Yes
End

Fig. 1 The diagram of EBA.

3 Our Proposed Method

In EBA, the movement of the artificial agents requires a constant, which is defined based on the chosen medium, to bind a range of the movable distance. Although EBA allows users to set the range in different values according to their demand, the diversity implanted into the artificial agents is still not strong and powerful for exploring the solution space. To jump out of the bottleneck caused by the constant used in EBA, we use the most common detectable periodic signal in the natural environment, which is the cosine signal, and a linear increasing dc signal to mixture the compound signal. Let $Iter_{tot}$ denotes the total iteration number, our mixture signal can be compounded by Eq. (4):

$$Wave(t) = A \cdot \cos(\omega t) + \frac{t}{Iter_{tot}} \tag{4}$$

where $Wave$ denotes the compounded mixture signal, A is the amplitude of the periodical signal, and t stands for the number of current iteration.

An example of the mixture signal is shown in Fig. 2.

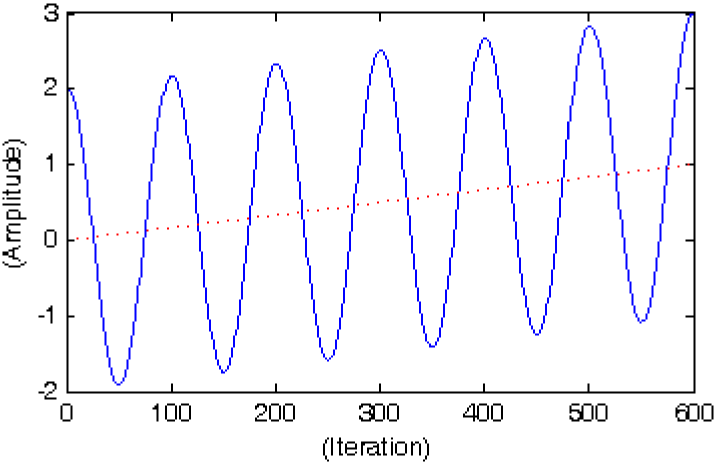


Fig. 2 Sample of the mixture signal in our proposed method.

As shown in Fig. 2, our proposed method utilize the mixture signal to increase the diversity of the searching range. The example given here uses $A = 2$ as the amplitude, and the length of a full cycle of the signal crosses 100 iterations.

To employ our proposed method in EBA, the process of the algorithm can be described as follows:

- Step 1. Initialization: Randomly spread the artificial agents into the solution space.
- Step 2. Movement: Change the solutions of the artificial agent by Eqs. (4)-(5). Generate a random number to decide whether takes the random walk process. If the random walk process is taken into the operation, Eq. (3) is employed.

$$x_i^t = x_i^{t-1} + Wave(t) \cdot \Delta T \tag{5}$$

- Step 3. Evaluation: Evaluate the fitness values of the artificial agents and update the near best solution found in the past iterations.
- Step 4. Termination: If the termination condition is satisfied, terminate the program and output the near best solution; otherwise, go back to step 2 and repeat the process.

4 Experiments and Experimental Results

To find out whether our proposed method provides positive contribution to the searching capacity of EBA, three test functions with known global optimum are used in the experiment. The test functions are listed as follows:

$$f_1(X) = \frac{1}{4000} \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \tag{6}$$

$$f_2(X) = 20 + e^1 - 20 \cdot e^{-0.2\sqrt{\sum_{i=1}^{n} \frac{x_i^2}{n}}} - e^{\sum_{i=1}^{n} \frac{\cos(2\pi \cdot x_i)}{n}} \tag{7}$$

$$f_3(X) = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 10\cos(2\pi \cdot x_i) + 10 \cdot n \tag{8}$$

In general, the feasible solutions to problems in engineering should be convergence. As the same as most engineering applications, the goal of the optimizations of all test functions in the experiments is to find the minimum value of the benchmark function. Moreover, since EBA is one of the algorithm in evolutionary computing, the way it finds solutions can be said as a kind of guided random search. Thus, all experiments are taken with 25 complete runs for finding the average results. The dimension of all test functions is set to 30; the total iteration number is set to 6000 for Eq. (6) and 5000 for Eqs. (7)-(8), and the initial range of these functions are set to [-600, 600], [3, 13], and [-5.12, 5.12]. The parameter setting of the original EBA and EBA with our proposed method is given in Table 1.

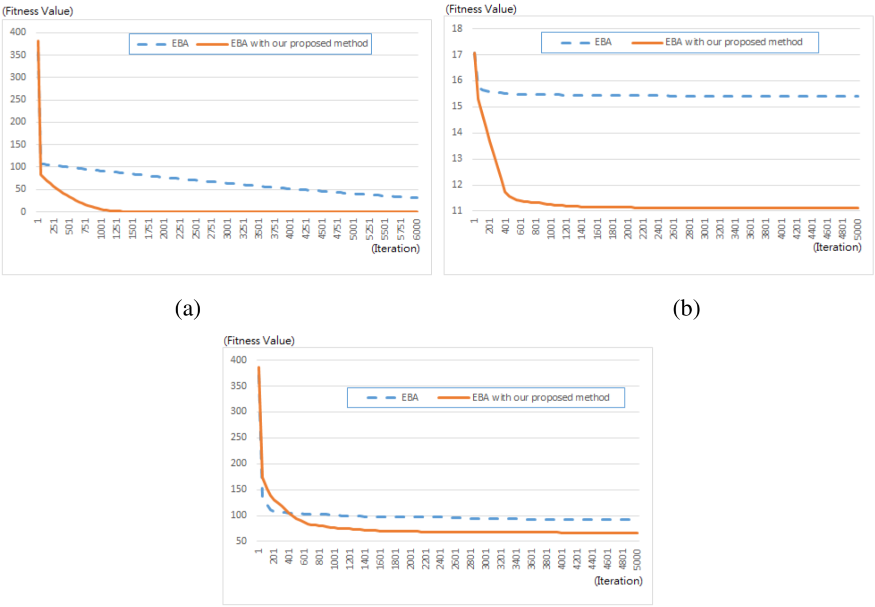
Table 1 Parameter setting for EBA and EBA with our proposed strategy.

	EBA [6]	EBA improved by our posed method
Fixed *r*	0.5	0.5
ΔT	[-1, 1]	[-1, 1]
A	NA	2

The experiments are taken with Matlab R2011b on a 64 bits Windows 8.1 operating system. Our experiments not only test the convergence speed, but also test for the accuracy on finding the near best solutions. To survey the accuracy, the final fitness values of all test functions are listed in Table 2 for compare. The final fitness value revealed in Table 2 indicates that our method improves the accuracy of EBA on finding the near best solution for all test functions in different level.

Table 2 The average results over 25 runs.

	EBA [6]	EBA improved by our posed method
$f_1(X)$	3.3647×10^0	2.0276×10^{-2}
$f_2(X)$	1.5395×10^1	1.1111×10^1
$f_3(X)$	7.9495×10^1	6.6817×10^1



(a) (b)

(c)

Fig. 3 The fitness values of all test functions: (a) for $f_1(X)$, (b) for $f_2(X)$, (c) for $f_3(X)$.

The convergence of the fitness values of all test functions is shown in Fig. 3. It is obvious that EBA with our proposed method presents faster convergence and higher accuracy in finding the near best solutions than the original EBA.

5 Conclusions

In this paper, EBA is improved by replacing the fixed constant used in the movement of the artificial agents by a continuously level-increased mixture signal. The signal is compounded of a periodical signal and a linear raising dc signal. The experimental results indicate that our proposed method improves the searching accuracy of EBA about 47.74 percent in average.

Acknowledgement The authors would like to thank Fujian University of Technology in China, for the financial support of this research.

References

1. Tsai, P.-W., Pan, J.-S., Liao, B.-Y., Tsai, M.-J., Vaci, I.: Bat Algorithm Inspired Algorithm for Solving Numerical Optimization Problems. Applied Mechanics and Materials **148–149**, 134–137 (2012)
2. Yang, X.-S.: A new metaheuristic bat-inspired algorithm. In: González, J.R., Pelta, D.A., Cruz, C., Terrazas, G., Krasnogor, N. (eds.) NICSO 2010. SCI, vol. 284, pp. 65–74. Springer, Heidelberg (2010)
3. Tsai, P.-W., Chen, C.-W.: A Novel Criterion for Nonlinear Time-Delay Systems Using LMI Fuzzy Lyapunov Method. Applied Soft Computing **25**, 461–472 (2014)
4. Hasançebi, O., Teke, T., Pekcan, O.: A bat-inspired algorithm for structural optimization. Computers and Structures **128**, 77–90 (2013)
5. Niknam, T., Azizipanah-Abarghooee, R., Zare, M., Bahmani-Firouzi, B.: Reserve Constrained Dynamic Environmental/Economic Dispatch: A New Multiobjective Self-Adaptive Learning Bat Algorithm. IEEE Systems Journal **7**(4), 763–776 (2013)
6. Tsai, P.-W., Pan, J.-S., Liao, B.-Y., Chu, S.-C.: Ehanced Artificial Bee Colony Optimization. International Journal of Innovative Computing Information and Control **5**(12(B)), 5081–5092 (2009)
7. Tsai, P.-W., Khan, M.K., Pan, J.-S., Liao, B.-Y.: Interactive Artificial Bee Colony Supported Passive Continuous Authentication System. IEEE Systems Journal **PP**, 1–11 (2012)
8. Chu, S.-C., Tsai, P.-W.: Computational Intelligence Based on the Behavior of Cats. International Journal of Innovative Computing, Information and Control **3**(1), 163–173 (2007)
9. Chu, S.-C., Tsai, P.-w., Pan, J.-S.: Cat swarm optimization. In: Yang, Q., Webb, G. (eds.) PRICAI 2006. LNCS (LNAI), vol. 4099, pp. 854–858. Springer, Heidelberg (2006)
10. Tsai, P.-W., Pan, J.-S., Chen, S.-M., Liao, B.-Y.: Enhanced Parallel Cat Swarm Optimization Based on the Taguchi Method. Expert Systems with Applications **39**(7), 6309–6319 (2012)

11. Temel, S., Unaldi, N., Kaynak, O.: On Deployment of Wireless Sensors on 3-D Terrains to Maximize Sensing Coverage by Utilizing Cat Swarm Optimization with Wavelet Transform. IEEE Transactions on Systems, Man, and Cybernetics: Systems **44**(1), 111–120 (2014)
12. Kong, L., Pan, J.-S., Tsai, P.-W., Vaclav, S., Ho, J.-H.: A Balanced Power Consumption Algorithm Based on Enhanced Parallel Cat Swarm Optimization for Wireless Sensor Network. International Journal of Distributed Sensor Networks **2015**, 1–10 (2015)
13. Pappula, L., Ghosh, D.: Linear antenna array synthesis using cat swarm optimization. International Journal of Electronics and Communications (AEÜ) **68**, 540–549 (2014)
14. Yang, F., Ding, M., Zhang, X., Hou, W., Zhong, C.: Non-rigid multi-modal medical image registration by combining L-BFGS-B with cat swarm optimization. Information Sciences (in press) (2015)

A New Digital Image Compression Method Using Chrestenson Transform

Mingyong Zhou, Zhongkan Liu and Hiromitsu Hama

Abstract In this paper, a new and innovative method to compress a digital image is proposed. This method first explores Chrestenson transform, then proposes p-adic process concept, and finally comes out with a practical way to form the p-adic process using real data, i.e., the digital image in this paper. It is first proved in [2] that Chrestenson transform of p-adic process is equivalent to K-L transform. We proposes a new method to generate the p-adic process from image data by extrapolating the data in a specific way. The result is verified and proved by initial simulations, and our method is shown to outperform the usual DCT transform.

1 Introduction

Discrete Cosine Transform(DCT) is well known to be close to Karhunen-Loeve (K-L) optimal transform for 1st order Markov stochastic process, thus DCT transform is widely used in standard digital image compression for long time including MPEG and JMEG standards. In 1955, H.F.Chrestenson proposed a generalized Walsh transform and extended Walsh transform into a more generic scenario for any integer number[1].In 1980s and 1990s, it is proved in[2][3] that for any p-adic process , the Chrestenson transform is equivalent to K-L transform in theory. However it is not shown in [2] how to construct a p-adic in a practical way, or find a process that is close to p-adic process. In this paper, we proposes a method to construct p-adic process by extrapolating the data, for example a digital image data, in a specific way so that a p-adic process can be constructed. It is shown

---

M. Zhou(✉)
Beihai College, Beihang University, Beijing, China
e-mail: zed6641@hotmail.com

Z. Liu
Beihang University, Beijing, China

H. Hama
Osaka City University, Osaka, Japan

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_4

that the method to construct the p-adic process is quite simple to one's surprise. It is shown in digital image compression simulations that all non-zero coefficient after Chrestenson transform is concentrated on 1-2 coefficients and the result is better than that of DCT.

2 Preliminary

Definition 1:(p-adic multiplication) for any integer p≥2, for any $x\in[0,+\infty)$, $t\in[0,+\infty)$, their p-adic integers can be represented as:

$$x = \left(x_{-s} \ldots x_{-2}x_{-1}x_0 \bullet x_1x_2 \ldots x_{r+1}\right)$$

$$t = \left(t_{-r} \ldots t_{-2}t_{-1}t_0 \bullet t_1t_2 \ldots t_{s+1}\right)$$

p-adic multiplication of t and x is defined as

$$t \otimes_p x = \sum_{k=-s}^{r+1} t_{1-k}x_k \tag{1}$$

where $\sum_{k=-s}^{r+1}$ denotes mod p sum.

Definition 2: (Chrestenson forward transform)

$$C(k,\omega) \underline{\underline{\Delta}} \exp\left\{-\frac{2\pi j}{p}\omega\otimes_p k\right\} \quad , \qquad k = 0,1,2\cdots\infty \tag{2}$$

is defined as Chrestenson function. For a series $x(n),(n = 0,1,\infty)$, its Chrestenson forward transform is defined as

$$X(\omega) \underline{\underline{\Delta}} \sum_{n=0}^{\infty} x(n)\exp\left\{-\frac{2\pi j}{p}\omega\otimes_p n\right\} \ , \ \omega\in[0,1) \tag{3}$$

Where p is an integer number larger or equal to 2. Note that in [2] a prime number is required, and in fact in reference [1] it is shown that p can be any integers larger than 2. $\otimes_p$ denotes a p-adic multiplication defined in Definition 1.

Theorem 1: Similar to Fourier theory, it can be verified that the Chrestenson inverse transform is as follows [2]

$$x(n) = \int_0^1 X(\omega) \exp\left\{ \frac{2\pi j}{p} \omega \otimes_p n \right\} d\omega \quad , \quad n = 0,1,2 \cdots \infty \tag{4}$$

The proof is omitted here as it can be refered in [2].

In fact, because of the characteristic of $\otimes_p$ operation, it can be easily verified that in theorem 1, the inverse transform result is discrete and is a step function. For p which is large, a fast algorithm can be proposed similar to FFT, but the details are not included in this paper.

Definition 3: (P-adic process) {x(n)} is defined as a p-adic stochastic process if auto-correlation $r(i) = E\{x(k)\overline{x(l)}\}$ depends only upon k -p l , where -p denotes mod p substraction.

For example, a simple auto-correlation matrix of p-adic stochastic process is as follows:

$$A = \begin{bmatrix} a_0 & a_1 + jb_1 & 0 & \dots & 0 & a_1 - jb_1 \\ a_1 - jb_1 & a_0 & a_1 + jb_1 & \dots & 0 & 0 \\ \dots & \dots & & \dots & & \\ \dots & \dots & & \dots & a_0 & a_1 + jb_1 \\ a_1 + jb_1 & 0 & 0 & \dots & a_1 - jb_1 & a_0 \end{bmatrix} \tag{5}$$

Where $r(0) = a_0, r(1) = a_1 + jb_1, r(2) = 0, r(3) = 0 \cdots$

Theorem 2: if $\xi(t)$ is a p-adic stochastic process, and its Chrestenson forward transform is defined as follows:

$$\xi_k = \frac{1}{p^n} \int_0^{p^n} \xi(t) w\left(k, \frac{t}{p^n}\right) dt, (k = 0,1,2, \cdots p^n - 1)$$

$$where \ , w\left(k, \frac{t}{p^n}\right) = \exp\{ -\frac{2\pi j}{p} (k \otimes_p t / p^n)\} \tag{6}$$

then

$$E\{\xi_k \xi_j\} = \delta_{kj} \frac{1}{p^n} \int_0^{p^n} r(\tau) w\left(k, \frac{\tau}{p^n}\right) d\tau,$$

$$(k, j = 0,1,2, \cdots p^n - 1) \tag{7}$$

Where $r(\tau)$ is the auto-correlation of $\xi(t)$. The proof can be found in reference [2].

Theorem 2 actually shows that for a p-adic stochastic process, it Chrestenson forward transform is equivalent to K-L transform. However in [2] it is not shown how to construct a p-adic process and in the next section we show a method to construct a p-adic process in a simple way.

3 Construct p-Adic Process by Extrapolation Method

For any integer p which is larger than 2, for convenience in this paper we assume $p = 2^m$, for example $p = 2^{10} = 1024$, for a real data x(1),x(2),…x(p/2), if the data are finite correlated, that is, r(l)=0, for certain l<p/2, it can be easily verified that the "image" extrapolation of data x(1),x(2),…x(p/2) by figure 1 method constructs a p-adic process , that is, the auto-correlation matrix after "image" extrapolation satisfies matrix A in Eq(5) .

[1]Figure 1 demonstrates a way how to "image" extrapolate the real data into an infinite data which construct a p-adic process.

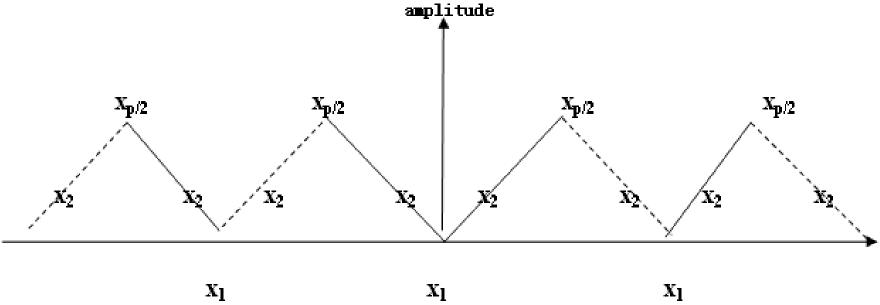


Fig. 1 "Image" extrapolation of data $x_1, x_2 \ldots x_{p/2}$ into infinite series.

4 Simulations of Digital Image Compression

In the simulation, we assume p=1024, a LENA picture of Figure 2 that contains 512 X 512 pixels is used for the simulation using MATLAB. To reduce the computation burden, a resize is manipulated, that is 64 X 8 =512 pixels are used. Every 64 pixels of each row are selected, and totally 8 lines are selected.

---

[1] Note: different from in Reference [2], data samples number in this paper is equivalent to integer p in Chrestenson function instead of p^m as proposed in [2]. This could be seen as a simplified scenario and could contribute to an easy construction of p-adic process as proposed in Figure 1.

The obtained 512 data are “image” extrapolated by using the method in the above section, that is, to generate another 512 “image” data which is totally 1024 data. The 1024 data are finally applied into the Chrestenson forward transform to obtain all 1024 coefficients. The simulation result is shown in Figure 3. Only the first two coefficients are non-zero while others are all zero, that is, (255,133,0,…… …0).

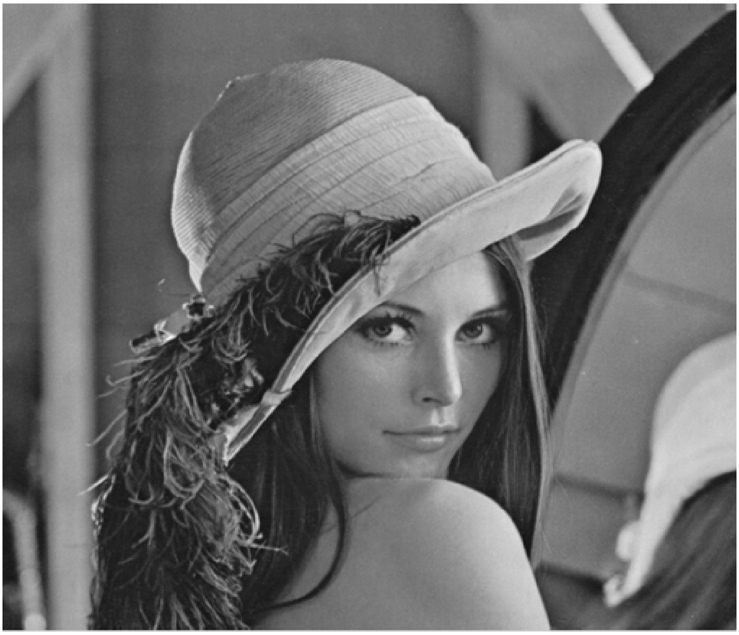


Fig. 2 LENA original picture 512x512 pixels

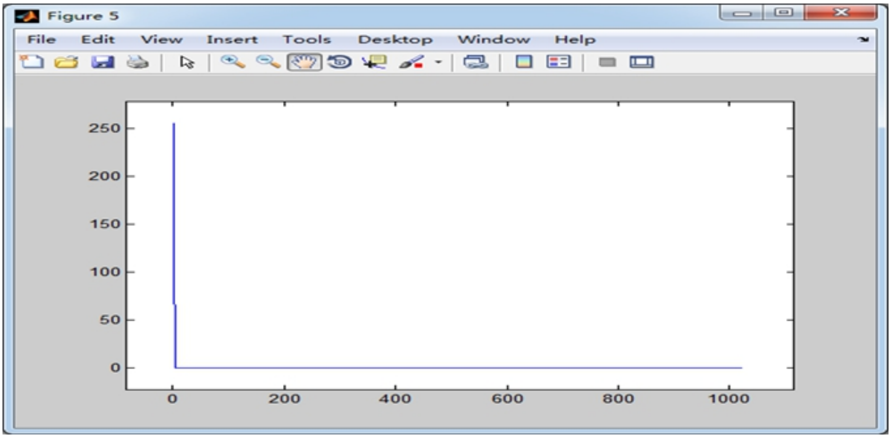


Fig. 3 All 1024 coefficients obtained from Chrestenson forward transform

5 Conclusion

A new image compression method is proposed in this paper. It is based on Chrestenson transform, which is a generic Walsh transform. A p-adic process is introduced and in this paper a simple and practical method to construct a p-adic process is proposed as well. Simulation shows that the coefficients obtained from Chrestenson forward transform concentrates on 1-2 coefficients, which verifies the theorem 2 for p-adic process. Initial simulation shows that it could be better than DCT transform that is widely used nowdays. For very large p integers, a fast algorithm can be introduced but it is not included in this paper.

References

1. Chrestenson, H.F.: A Class of gerneralized Walsh functions. Pacific Journal of Math. **5**, 17–31 (1955)
2. Zhongkan, L.: Chrestenson transform representation theorems and its applications. Journal of Beihang University, **2** (1987) (in Chinese)
3. Zhou, M., Liu, Z., Hama, H.: A Resolution-Controllable Harmonical Retrieval Approach on the Chrestenson Discrete Space. IEEE transactions on Signal Processing, **42**(5), May 1994

Intelligent Algorithms and Their Application in Electrical Impedance Tomography

Xin Lu and Mingyong Zhou

Abstract EIT (Electrical impedance tomography) problem is an inverse problem of mathematical physics. In the past years, numerous scientists and researchers put their effort to this problem and a lot of traditional mathematics algorithms have been applied. The difficulty of this problem is that we do not know inside information, all we know is potential on the surface of the object, then use this known data to calculate inside structure. With the time going by, Intelligent algorithms have become more popular. In this paper, we used an intelligent algorithm Back Propagation (BP) neural network algorithm to solve this problem instead of complex traditional mathematics algorithm. Different from normal usage of BP neural network algorithm, we look each nodes in different layers as the each result in solution this time. Mend each node weight constantly by some groups of prior data, we will get the solution which is consist of each node weight. In the experiment simulation, solutions of iteration with 500times and 1000times show that this algorithm is feasible. At the end of the paper, we concluded this experiment and analysis the executable of another intelligent algorithm which will be used in this EIT problem, Genetic algorithm.

Keywords EIT · Inverse problem · Intelligent algorithm · Neural network

1 Introduction

EIT (Electrical impedance tomography) is a newly and burgeoning image reconstruction technology [1]. It is required that we need to set a circle of electrodes on the surface of the object, then we can reconstruct the interior image when current being injected from the electrode and receiving from other electrodes. Compared

---

X. Lu(✉) · M. Zhou
Beihai College of Beihang University, Beihai City 536000, Guangxi, China
e-mail: luxin0311@live.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_5

with other image reconstruction technology like CT and MRI, the advantage of EIT is budget and non-invasive. Intelligent algorithms take an important role in computing now, as one of the most popular intelligent algorithms, neural network algorithm is based on logical inference, which is often rely on computers. According to it's requirement, we need to construct the structure of neural network system first, then use prior data to train the weight in each node. Finally, we can use that system when the training is finished. The distinct feature of it is that it does not need complex mathematical computation, once upon we constructed the system model, the calculating process will be both easy and fast.

2 Basic Model of EIT Problem

As the technology of EIT researched in 70s, various kind of image reconstruction algorithms have been used [2-4].

Inside of magnetic field, magnetic field effect can be ignored due to the quite low permeability in biological tissue. According to the MaxWell theory and Ohm law, we can get the description like this:

$$-\rho^{-1}\nabla\phi=\vec{J} \tag{1-1}$$

In this equation, ρ represents the distribution function of resistivity, ϕ means potential distribution, J is the function of boundary electric current density.

Due to there is no electric current through interior of biological tissue, so we get this:

$$\nabla\cdot\vec{J}=0 \tag{1-2}$$

Combined (1-1) and (1-2), we can get equation:

$$\nabla\cdot\rho^{-1}\nabla\phi=0 \tag{1-3}$$

This partial differential equation should satisfy the Dirichlet boundary condition

$$\phi\,|_{\partial\Omega}=V \tag{1-4}$$

In that, **Ω** is the area where the object in. For the electrodes with inject current, Neumann boundary condition is:

$$\left.\rho^{-1}\frac{\partial\phi}{\partial\vec{n}}\right|=\vec{J} \tag{1-5}$$

The equations above constitute the mathematical model of EIT problem [5].

3 Basic Back Propagation(BP) Neural Network Algorithm Inference

Basic Back propagation includes two parts, one is the forward propagation of signals the other part is the backward propagation of errors. That is to say, get result from forward direction and mend the weights and thresholds from backward direction [6].

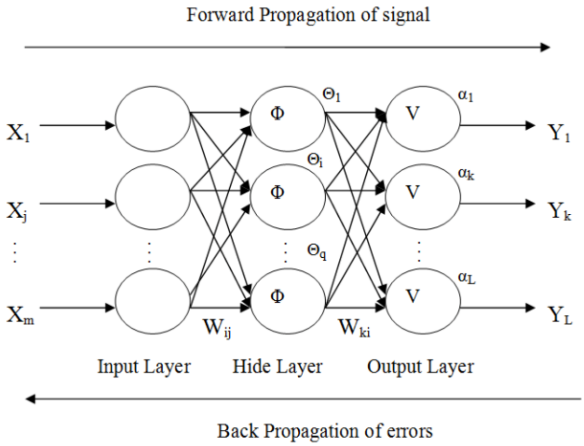

Forward Propagation of signal
X_1
X_j
X_m
Φ
Θ_1
Θ_i
Θ_q
V
α_1
α_k
α_L
Y_1
Y_k
Y_L
W_{ij}
W_{ki}
Input Layer
Hide Layer
Output Layer
Back Propagation of errors

Fig. 1 The back propagation neural network model.

In this chart
Xj represents No.j node input in Input Layer, and the range of j is [1,m].
Φ(x) represents stimulate function of Hide Layer.
V(x) represents stimulate function of Output Layer.
Wij represents the weight between No.i node in Hide Layer and No.j node in Input Layer.
Θi represents the threshold of No.j node in Hide Layer.
Wki represents the weight between No.k node in Output Layer and No.i node in Hide Layer.
αk represents the threshold of No.k node in Output Layer.
Yk represents the output of No.k node in Output Layer.
(1). The process of forward propagation of signal
The input of No.i node in Hide Layer, Hi:

$$H_i = \sum_{j=1}^{m} w_{ij} x_i + \theta_i \qquad (2\text{-}1)$$

The output of No.i node in Hide Layer Yi:

$$Y_i = \Phi(\sum_{j=1}^{q} w_{ij} x_j + \theta_i) \tag{2-2}$$

The input of No.k node in Output Layer

$$H_k = \sum_{i=1}^{q} w_{ki} \Phi(\sum_{j=1}^{m} w_{ij} x_j + \theta_i) + \alpha_k \tag{2-3}$$

The output of No.k node in Output Layer

$$Y_k = V\sum_{i=1}^{q} w_{ki} \Phi(\sum_{j=1}^{m} w_{ij} x_j + \theta_i) + \alpha_k \tag{2-4}$$

(2) The process of Back Propagation of errors

The process of Back Propagation of errors is means that the first time errors will be given from Output Layer, then be given back Layer by Layer. Then, mend the node weight and threshold through steepest descent algorithm.

For the sample data P to say, there is a error function

$$O_p = \frac{1}{2}\sum_{k}^{L} (T_k - Y_k)^2 \tag{2-5}$$

Based on the steepest descent algorithm of errors, we mend each node weight and threshold constantly. We give each correction as follows:

Formula for the node weight in Output Layer

$$\Delta w_{ki} = -\eta \frac{\partial O}{w_{ki}} = -\eta \frac{\partial O}{\partial Y_k} \frac{\partial Y_k}{\partial H_k} \frac{\partial H_k}{\partial w_{ki}} \tag{2-6}$$

Formula for the threshold in Output Layer

$$\Delta \alpha_k = -\eta \frac{\partial O}{\alpha_k} = -\eta \frac{\partial O}{\partial \alpha_k} \frac{\partial Y_k}{\partial H_k} \frac{\partial H_k}{\partial \alpha_k} \tag{2-7}$$

Formula for the node weight in Hide Layer

$$\Delta w_{ij} = -\eta \frac{\partial O}{w_{ij}} = -\eta \frac{\partial O}{\partial \varphi_i} \frac{\partial \varphi_i}{\partial H_i} \frac{\partial H_i}{\partial w_{ij}} \tag{2-8}$$

Formula for the threshold in Hide Layer

$$\Delta \theta_i = -\eta \frac{\partial O}{\theta_i} = -\eta \frac{\partial O}{\partial \varphi_i} \frac{\partial \varphi_i}{\partial H_i} \frac{\partial H_i}{\partial \theta_i} \tag{2-9}$$

As a conclusion, we get these:

$$\Delta w_{ki} = \eta \sum_{p=1}^{p} \sum_{k=1}^{l} (T_k^p - Y_k^p) \cdot V' H_k \cdot \varphi_i \tag{2-10}$$

$$\Delta \alpha_k = \eta \sum_{p=1}^{p} \sum_{k=1}^{l} (T_k^p - Y_k^p) \cdot V' H_k \tag{2-11}$$

$$\Delta w_{ij} = \eta \sum_{p=1}^{p} \sum_{k=1}^{l} (T_k^p - Y_k^p) \cdot V' H_k w_{ki} \cdot \Phi' H_i \cdot x_j \tag{2-12}$$

$$\Delta \theta_i = \eta \sum_{p=1}^{p} \sum_{k=1}^{l} (T_k^p - Y_k^p) \cdot V' H_k w_{ki} \cdot \Phi' H_i \tag{2-13}$$

4 Inverse BP Neural Network on EIT Model Building

Consider about the EIT problem is a inverse problem, we can not solve this kind of problem directly. Thus, we can not use neural network directly like we using it in other questions to solve this inverse problem. To deal with this, we present another way to use neural network to solve that. We assume that the function of impedance is consist of uniform distribution nodes, which will be our target solution.

Normally, before we use neural network system, we use some groups of data to train for suitable nodes. Then, we use this trained system to solve the problem what we want to. But this time, we do it in a inverse way. We assume the nodes of neural network system to be the result of the EIT problem. We construct a m×n matrix like this:

$$\left|\begin{matrix} \Omega_{11}\Omega_{12}\Omega_{13}\cdots\Omega_{1m} \\ \Omega_{21}\Omega_{22}\Omega_{23}\cdots\Omega_{2m} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \Omega_{n1}\Omega_{n2}\Omega_{n3}\cdots\Omega_{nm} \end{matrix}\right|$$

Fig. 2 EIT structure matrix.

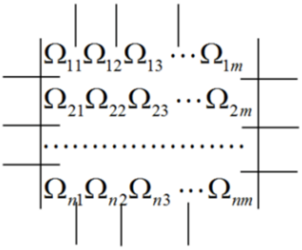

$\Omega_{11}\Omega_{12}\Omega_{13}\cdots\Omega_{1m}$
$\Omega_{21}\Omega_{22}\Omega_{23}\cdots\Omega_{2m}$
$\Omega_{n1}\Omega_{n2}\Omega_{n3}\cdots\Omega_{nm}$

Fig. 3 EIT structure matrix(with current nodes).

Each element in this matrix is a node we assumed in neural network system. Given to it is a circle and always only one input current point the others are receiving current, it can be constructed as figure 3 (each point could be inputted current).

5 Simulation Experiment

In order to have a real data to compared with and analysis, we use some resistances to build a circuit. Before we using software (Matlab 7.0) to simulate, we measured and recorded the current from each point. Here is the real circuit with real data. Current input is from the point with '+', outflow point is marked with '-':

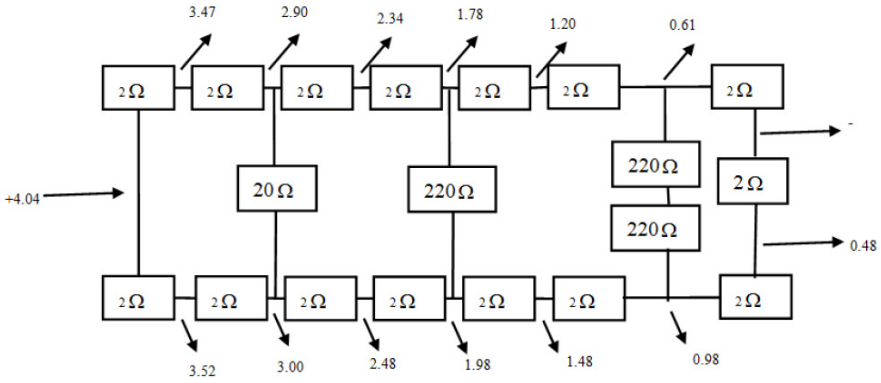

3.47
2.90
2.34
1.78
1.20
0.61
2Ω
2Ω
2Ω
2Ω
2Ω
2Ω
2Ω
-
+4.04
20Ω
220Ω
220Ω
2Ω
220Ω
0.48
2Ω
2Ω
2Ω
2Ω
2Ω
2Ω
2Ω
3.52
3.00
2.48
1.98
1.48
0.98

Fig. 4 The real tested model structure (with all known information).

As we know, in the real application, we never will know the inside structure. So, we build a 4*5 even matrix like this:

$$\rho = \begin{vmatrix} \Omega_{11} & \Omega_{21} & \Omega_{31} & \Omega_{41} & \Omega_{51} \\ \Omega_{12} & \Omega_{22} & \Omega_{32} & \Omega_{42} & \Omega_{52} \\ \Omega_{13} & \Omega_{23} & \Omega_{33} & \Omega_{43} & \Omega_{53} \\ \Omega_{14} & \Omega_{24} & \Omega_{34} & \Omega_{44} & \Omega_{54} \end{vmatrix}$$

Fig. 5 The model of ρ.

What we can know and only know in real is potential, which is our prior data:

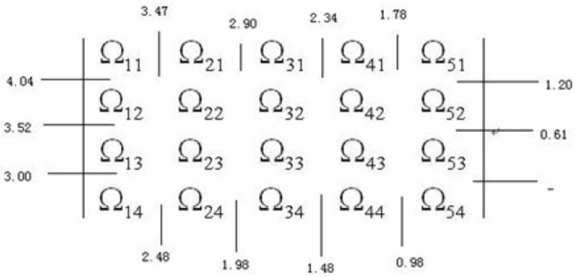

3.47
2.90
2.34
1.78
4.04
3.52
3.00
1.20
0.61
-
Ω_{11} Ω_{21} Ω_{31} Ω_{41} Ω_{51}
Ω_{12} Ω_{22} Ω_{32} Ω_{42} Ω_{52}
Ω_{13} Ω_{23} Ω_{33} Ω_{43} Ω_{53}
Ω_{14} Ω_{24} Ω_{34} Ω_{44} Ω_{54}
2.48
1.98
1.48
0.98

Fig. 6 The matrix with current that we can know.

Before use BP neural network algorithm, we need to initialize all the node weight, in here, we initialize the node weight as this:

$$\rho_0 = \begin{vmatrix} 10 & 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 & 10 \end{vmatrix}$$

Fig. 7 Initialize the original matrix.

The BP neural network algorithm we used here is use the initialize node data as the real data to calculate the potential, then use the difference between this potential from simulation and the real potential to mend the node weight. When this has been done, go to the next iteration, until the solution of this problem is gratifying. We illustrate after 500 times and 1000 times iteration solution here:

$$\rho_{500} = \begin{vmatrix} 3.02 & 3.01 & 7.5 & 8.33 & 3.18 \\ 2.45 & 16.92 & 711.13 & 113.33 & 32.81 \\ 3.54 & 691.12 & 89.65 & 599.65 & 25.75 \\ 2.59 & 2.77 & 7.56 & 23.33 & 19.63 \end{vmatrix}$$

Fig. 8 The matrix after 500 times iteration of BP.

$$\rho_{1000} = \begin{vmatrix} 3.71 & 2.96 & 4.16 & 5.36 & 4.09 \\ 2.08 & 18.89 & 693.23 & 125.78 & 27.36 \\ 3.69 & 783.23 & 88.44 & 652.39 & 17.32 \\ 3.02 & 3.85 & 6.57 & 17.36 & 18.59 \end{vmatrix}$$

Fig. 9 The matrix after 1000 times iteration of BP.

6 Constructed by Genetic Algorithm

In genetic algorithm each solution of problem is called chromosome, which means each individual in the group. Chromosome is a code string coded in a given method, each code unit in the code string is called gene. Fitness value can tell the quality of each chromosome, fitness value is calculated by evaluation function[7].

To deal with the EIT problem, we need to find the mapping relationship between EIT model and genetic algorithm. In here we still use $\boldsymbol{\rho}$ to represent the solution of EIT problem. Same as mentioned in BP neural network algorithm, we need to initialize the solution $\boldsymbol{\rho}$ in the first step. But in here, we initialize four solutions because chromosomes will mate and evolve. Then, we need a evaluation function to give the fitness value of each condition of solutions. This evaluation function is based on the function of variance of each point at the edge of object. After that, these chromosomes will be chosen and go mating, variant and evolution. Finally, judge the solution whether it is satisfied our requirement. If not, back to be chosen again.

7 Conclusion

Through Back Propagation (BP) neural network algorithm, we can easily find that the whole calculation process is in a simple mathematics model without complex operation, which is the prominent feature of intelligent algorithm. From the solution what we have solved, we can see the solution become more steady without obvious change in each node. Through this experiment, back propagation neural network algorithm can be used to solve this inverse mathematics problem has been proved useful. Lastly, talked about another intelligent algorithm for EIT problem which is genetic algorithm and it is feasible too.

References

1. Xiuzhen, D.: The situation and challenges of EIT problem. Chinese Journal of Biomedical Engineering **27**(5), 64–646 (2008)
2. Weberster, J.G.: Electrical Impedance Tomography, pp. 97–137. Adam Hilger, Bristol (1990)
3. Yorkdey, T.J., et al.: Computer reconstruction algorithm for electrical impedance tomography. IEEE Trans. Biomed. Eng., BME **34**(11), 843–852 (1987)
4. Woo, E.J., et al.: A Robust image reconstruction algorithm for electrical impedance tomography. IEEE Trans. Med. Imaging, MI **12**(2), 137–146 (1993)
5. Yan, T., Jikuan, C., Zhongkan, L.: Use combined variable-metric method to solve EIT problem. Chinese Journal of Biomedical Engineering **2**, 18–20 (1997)
6. Jun, Z., Zhijun, Z., et al.: Computational Algorithm, pp. 13–26. Tsinghua University Press, Beijing (2009)
7. Holland, J.H.: Genetic algorithm. Scientific American, 44–50 (1992)

Design of Multifunctional Communication Card Based on Reflective Memory

Shou-Da Jiang, Jia-Yu Zhao and Chang-An Wei

Abstract According to the requirement of real-time performance and universality in the virtual simulation system, the communication card adopts reflective memory technology with the ability to build the reflective memory network. At the same time, according to the requirement of interaction between virtual model and physical model, the card has serial interface and the ability of interaction with reflective memory network through serial interface. The experiments proved that the multifunctional communication card's bandwidth is up to 30 MB/s, its bit error rate is lower than 10^{-15}, its time delay is up to the rank of the deep microsecond.

Keywords Reflective memory · Multifunctional communication · Virtual simulation

1 Introduction

Along with the improvement of the equipment's performance, the control cycle has already reached 1ms even the rank of microseconds. Therefore, the reflective memory network is usually applied in semi-physical simulation system for data transmission [1]. The semi-physical simulation system of armament generally consists of parts, which is replaced by virtual models partly or completely. The virtual models transfer data through the network. The virtual models and the physical models transfer data through the serial interface, such as RS422 and 1553B. As a result, the semi-physical simulation system has the requirement of the direct data transmission between the data of reflective memory network and the serial interface.

S.-D. Jiang · J.-Y. Zhao · C.-A. Wei(✉)
Department of Automatic Testing and Control, Harbin Institute of Technology,
No.2 Yi-Kuang Street, Nangang District, Harbin 150080, China
e-mail: {zhao4471437,nathanwei}@163.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_6

At present, reflective memory network products on the market can't meet the requirements of semi-physical simulation system for network function. So the development of multifunctional communication module based on reflective memory is very significant for the semi-physical simulation system.

2 Principle and Working Mode of Multifunctional Communication Card

Reflective memory network, which is composed of the multifunctional communication card, is connected by optical fiber [2]. Every computer online can form a node on the network by inserting a multifunctional communication board on CPCI slot. Fig. 1 shows the virtual simulation system.

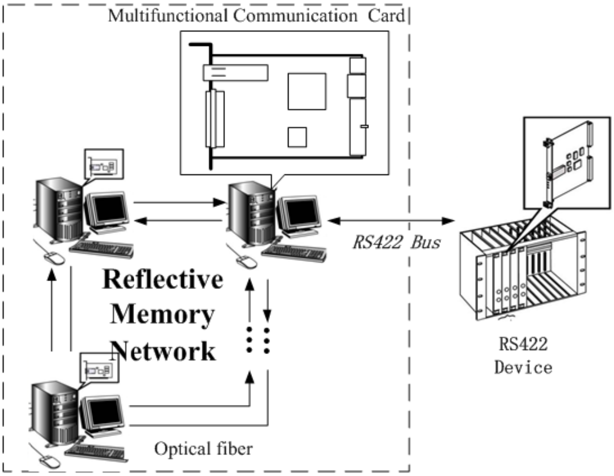


Fig. 1 Virtual simulation system structure

Each node on the network of global memory is mapped to a virtual local memory, which constitutes the distributed shared memory. The operation of reading and writing of the local node memory is equivalent to that of the global memory [3]. The local node memory can be read and written by the host machine or device with RS422 interface. Multifunctional communication interface card is integrated with 4-way RS422 interface, so that you can directly access to the reflective memory network by RS422 interface.

3 Overall Scheme Design

Generally, the communication card is composed of the host machine interaction module, the DDR2 module, RS422 module, and optical fiber interface module (Fig. 2).

Four modules mount on the Avalon bus, with the aid of the bus for data interaction. The on-line strategy defined by Avalon Switch-based bus makes any Avalon Bus master device communicate with any salve device [4].

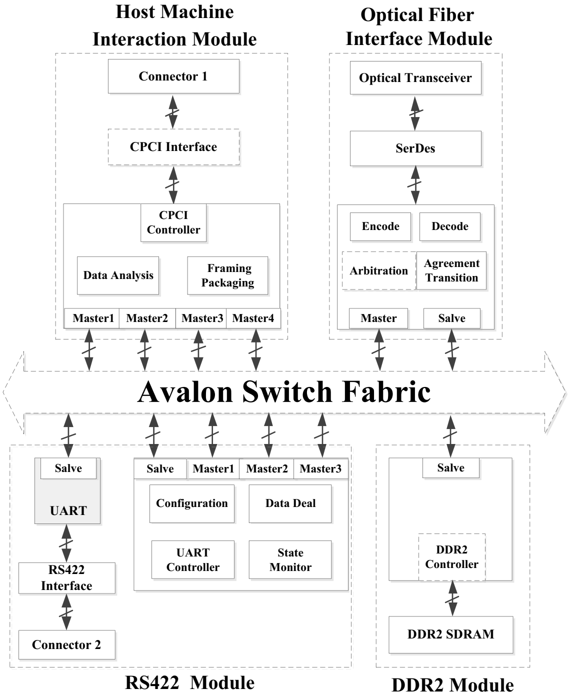

Host Machine Interaction Module
Connector 1
CPCI Interface
CPCI Controller
Data Analysis
Framing Packaging
Master1
Master2
Master3
Master4
Optical Fiber Interface Module
Optical Transceiver
SerDes
Encode
Decode
Arbitration
Agreement Transition
Master
Salve
Avalon Switch Fabric
Salve
UART
RS422 Interface
Connector 2
Salve
Master1
Master2
Master3
Configuration
Data Deal
UART Controller
State Monitor
Salve
DDR2 Controller
DDR2 SDRAM
RS422 Module
DDR2 Module

Fig. 2 Multi-function communications overall block diagram

3.1 *Design of Host Machine Interaction Module*

Host Machine interaction module is responsible for handling information interaction, implementing data parsing, data framing and packaging, and Avalon bus interaction. The module is composed of the CPCI-bus interface unit, data analytical unit, data framing and packaging unit, etc. (Fig. 3).

PLX9054 is on C operating mode. Its transmission mode is the DMA mode [5].

Data analytical unit distinguishes different types of data by PCI bus address. There are 3 kinds of data types: Data via RS422 interface, Data broadcast to the reflective memory network, Data that is read or written from local memory. In order to manage data better, the reflective memory is divided into two areas: the memory whose address is 0x00000000 ~ 0x000ffff storages the data of RS422; the memory whose address is 0x0010000~0x8000000 storages the data of general memory.

Data framing and packaging is to solve problems such as different data bandwidth, protocol mismatching. For the convenience of data interaction, data structures of all stripes are unified.

In order to extend the system's bandwidth, the host machine interaction module integrates four Avalon bus master ports. Since each bus host has its own dedicated Internet, bus host only needs to take share from the machine, rather than the bus itself. So that four Avalon bus master ports will not result in the bus congestion.

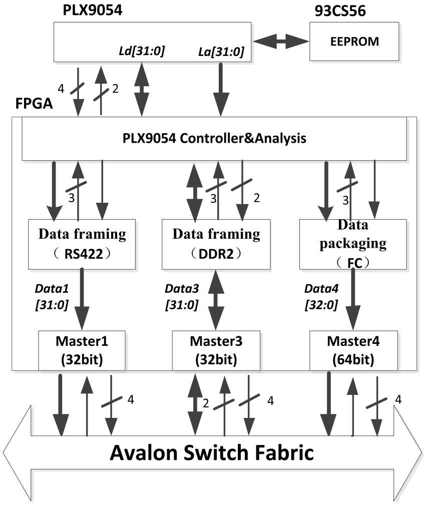

PLX9054
93CS56
Ld[31:0]
La[31:0]
EEPROM
4
2
FPGA
PLX9054 Controller&Analysis
3
3
2
3
Data framing（RS422）
Data framing（DDR2）
Data packaging（FC）
Data1 [31:0]
Data3 [31:0]
Data4 [32:0]
Master1 (32bit)
Master3 (32bit)
Master4 (64bit)
4
2
4
4
Avalon Switch Fabric

Fig. 3 The host machine interaction module block diagram

Avalon interface can reach 1 time per clock transmission [6], so the impact on the uplink and downlink data transmission rate can be neglected.

3.2 Design of Optical Fiber Interface Module

Optical fiber interface module is composed of optical transceiver unit, encoding and decoding unit, data of arbitration and data analysis unit, and so on(Fig. 4).

Optical transmission module adopts integrated photovoltaic solutions [7]. And the maximum serial data transfer rate reaches 1.0625 Gb/s [8].

In order to maintain data integrity among various functional units, multiple FIFOs are designed to cache data receiving and sending. The FIFOs solve the problems about the data rate mismatch and the clock domain mismatch.

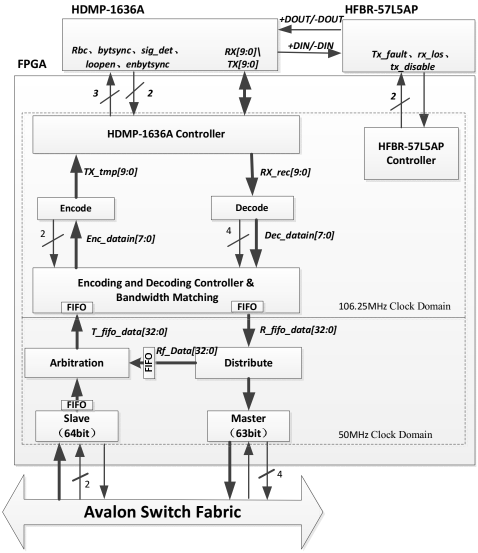

HDMP-1636A
HFBR-57L5AP
FPGA
+DOUT/-DOUT
+DIN/-DIN
Rbc、bytsync、sig_det、loopen、enbytsync
RX[9:0]\ TX[9:0]
Tx_fault、rx_los、tx_disable
3
2
2
HDMP-1636A Controller
HFBR-57L5AP Controller
TX_tmp[9:0]
RX_rec[9:0]
Encode
Decode
2
Enc_datain[7:0]
4
Dec_datain[7:0]
Encoding and Decoding Controller & Bandwidth Matching
FIFO
FIFO
106.25MHz Clock Domain
T_fifo_data[32:0]
R_fifo_data[32:0]
Arbitration
FIFO
Rf_Data[32:0]
Distribute
FIFO
Slave (64bit)
Master (63bit)
50MHz Clock Domain
2
4
Avalon Switch Fabric

Fig. 4 Optical fiber interface module block diagram

8b/10b decoder built in the high-speed serial transceiver can check out the single bit error, and at the same time CRC check can effectively find errors and correct single bit error [9]. Using these two schemes can effectively give feedback of processing error to the upper.

3.3 Design of RS422 Module

RS422 module is mainly composed of UART unit, control unit and RS422 interface circuit.

UART in Avalon bus system is a common character of peripheral equipment. UART provides the communication mode of serial character stream for embedded systems and external equipment on Altera FPGA. Control unit's main functions include UART controller, address matching, and data packaging, etc. Three Avalon master ports and one Avalon salve port are embedded into the Control unit. Through the Avalon bus, RS422 module can interact with other modules.

4 Function and Performance Test

The virtual simulation system consisting physical device and virtual device is set up to test functions and performance of the multi-function card based on reflective memory.

4.1 Function of Reflection Network Test

Test network has three nodes, an optical interconnection of annular links. One of the network nodes accesses to RS422 device. Fig. 5 shows the test model.

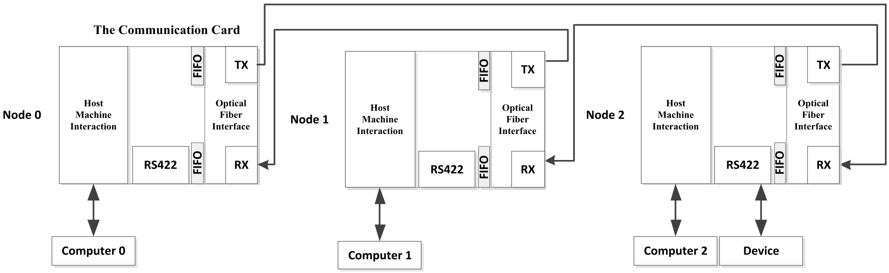

The Communication Card
Node 0
Host Machine Interaction
FIFO
TX
Optical Fiber Interface
RS422
FIFO
RX
Computer 0
Node 1
Host Machine Interaction
FIFO
TX
Optical Fiber Interface
RS422
FIFO
RX
Computer 1
Node 2
Host Machine Interaction
FIFO
TX
Optical Fiber Interface
RS422
FIFO
RX
Computer 2
Device

Fig. 5 Reflection network test model

In radio communication test, Node 0 as a data source sends packets to other nodes. After receiving the packet, Node 1, 2 execute store operation. At the same time, Node 1, 2 forward packets in sequence. When the packet returns to the Node 0, the data packets are deleted, no longer forwarding. Finally, we compare three nodes memory data. This test was conducted on the size of 12 bytes of data packets by using continuous testing for a long time. This test is a continuous testing using the data packets, the size of 12 bytes. Actual test results show that the transmission error rate is lower than 10^{-15}.

4.2 Function of RS422 Test

The RS422 device on Node 3 radios data packets as a reflective memory network data source. According to three kinds of baud rate: 9600 bps, 115.2 Kbps, 1 Mbps and word length: 8/9, each channel has a total of 6 kinds of combination mode test respectively. The data sent and received by RS422 device is exactly the same as data on the reflective memory.

4.3 Network Transmission Rate Test

In the rate test, Node 0 initiates a data transmission. At the same time the system starts 0.1ms clock provided by RTX system [10]. When the data update is completed, we stop recording. Finally, we can get the data writing rate. In the test, we write 100 data packets every time and we write 500 times. Total valid data is 200GB. Table 1 shows the result of testing of data writing.

Table 1 Testing of data writing

	Data packets length	Fiber length	Time-consuming	Writing rate
Test.1	4B	10m	6.36h	8.73MB/s
Test.2	32B	10m	2.15h	25.85MB/s
Test.3	64B	10m	1.88h	29.55MB/s

Then, we read memory data by hosting computer. We read 100 data packets every time and we read 500 times. Total valid data is 200GB. Table 2 shows the result of testing of data reading.

Table 2 Testing of data reading

	Data packets length	Fiber length	Time-consuming	Reading rate
Test.1	4B	10m	6.81h	8.15MB/s
Test.2	32B	10m	2.3h	24.14MB/s
Test.3	64B	10m	2.01h	27.58MB/s

We can draw conclusion as following that when the amount of data is small, the transmission rate is not high, because most of the time overhead is the interrupt transmission and logical judgment. When data gets bigger, the transmission rate is faster, and tends to be stable (30MB/s).

4.4 Transmission Delay Test

Delay in transmission includes the time delay between nodes and the time delay in the node. The test model is as follows:

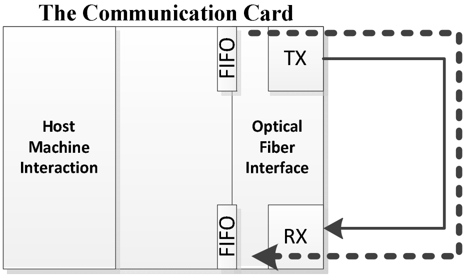


Fig. 6 Transmission delay test model

A data transmission process between adjacent nodes needs to go through units, such as sending unit, optical transceiver, optical fiber, optical transceiver, receiving unit. The length of optical fiber is 10 m.

Through 100 times test, delay between nodes can be measured with an average of 244.4 ns.

The internal processing of data in the node will cause the delay, The test model is as follows:

A data transmission process above includes 2 time delay between two nodes and 1 time delay in the node. Through 100 times test, delay in the node can be measured with an average of 263.2 ns.

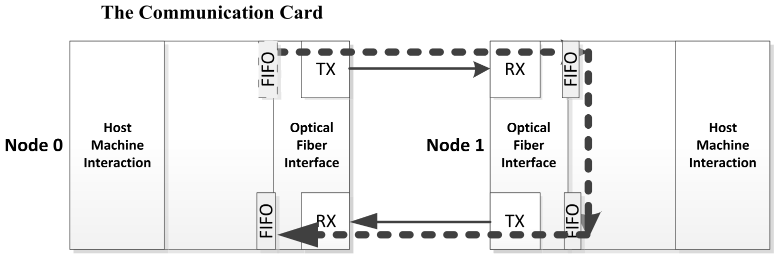


Fig. 7 Internal delay test model

5 Conclusions

A series of experimental tests show that the Multifunctional communication card based on reflective memory is designed reasonably. It satisfies many requirements such as high speed, reliability, predictability, multifunction, etc. At present, the card has been applied in virtual simulation platform. It should be noted that the board integrates RS232, 1533B, CAN and other communication modules, and the communication capabilities of the board can be extended easily by altering the programs in FPGA. The multifunctional communication card based on reflective memory is bound to have a good prospect.

Acknowledgement This work is supported by National Natural Science Foundation of Youth Science Foundation under Grant 61301207.

References

1. Chia, S., Ichiro, M.: RT-CRM: real -time channel-based reflective memory. IEEE Transactions on Computers **49**(11), 1202–1214 (2000)
2. VMIPCI-5565 Ultrahigh Speed Fiber-Optic Reflective Memory with Interrupts, GE Hardware Reference, pp. 26–57 (2008)
3. Bershad, B.N., Zekauskas, M.J., Sawdon, W.A.: Midway Distributed Shared Memory System, vol. 4, pp. 528–537. IEEE Computer Society Press, Los Alamitos (1993)
4. Qian, Z., Yu-Kun, S.: A design of muti-core system based on Avalon bus. In: Proceedings of International Conference on Computer Science and Network Technology (2011)

5. Bittner Ray. Bus mastering PCI express in an FPGA. In: The 7th ACM SIGDA International Symposium on Field-Programmable Gate Arrays, vol. 2, pp. 273–276 (2009)
6. PCI-5565PIORC* Ultrahigh Speed Fiber-Optic Reflective Memory with Interrupts. GE Hardware Reference
7. Gigabit Ethernet and Fiber Channel SerDes ICs Technical Data. Agilent Technologies (2002)
8. Jovanoie, M., Milutinovie, V.: An Overview of Reflective Memory Systems. IEEE Concurrency **7**(2), 56–64 (1999)
9. Robertson, I., Hetherington, G.: Testing high-speed, large scale implementation of SerDesI/Os on chips used in throughput computing systems. In: Proceedings International Test Conference, pp. 992–999 (2005)
10. You, T., Du, C., Zhu, Y.: Supporting technology for virtual numerical control system based on RTX and reflective memory network. In: The Fifth Natural Computation International Conference, vol. 8, pp. 319–323 (2009)

Computation of Large-Scale Electric Field in Free Space

Chang-An Wei, Qiao-Chu Cui, Lian-Lei Lin and Shou-Da Jiang

Abstract A method of modeling Spatial Electric Field around the earth is presented to provide the spatial electric field data in virtual test system. Firstly, the electric field in different time in the condition of sunny day is modeled to generate the electric field data in a period time and spatial. Then, this paper gives the analysis of the influence to Spatial Electric Field of thunderstorm.

Keywords Spatial electric field · Virtual test system · Cloud flash

1 Introduction

The spatial electric field is an important parameter of earth environment, and worthy of deeply study from the view of space physics basic theory and the application of aerospace engineering [1]. The atmosphere around the earth in the vertical direction can be divided into the troposphere, the stratosphere, the middle layer, heat layer and outer lay [2]. The troposphere is the lowest layer, and is in most close relations with human. The aircraft taking off and land in, is flying in the stratosphere [3]. The stratosphere has a lot to do with latitude, at the low-altitude area, at the latitude is above 16 to 18 km, at the mid-latitude area, the latitude is 10 to 12 km; at the high-latitude area, the latitude is 8 to 9 km. In this paper, the scope of space electric field is limited between 0 and 30 km.

Environmental modeling is a scientific and effective simulation method. As so far, there is no modeling method of the space electric field with the reason of it is influenced by all kinds of weather conditions in the atmosphere, such as thunderstorm and cloud flash, etc.

C.-A. Wei · Q.-C. Cui · L.-L. Lin(✉) · S.-D. Jiang
Department of Automatic Testing and Control, Harbin Institute of Technology,
No. 2 Yi-Kuang Street, Nangang District, Harbin 150080, China
e-mail: nathanwei@163.com, linlianlei@hit.edu.cn

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/ 978-3-319-23204-1_7

This paper presents a method to modelling the space electric field, and using the model, we can generate database for a specific time and space electric field, then the modeling precision is analyzed.

2 Model of Spatial Electric Field in Sunny Day

Using the following standard to determine the sunny days [4]:

(1) Local time between 09:00 am until 4 p.m. at least 90% of the time the sun point-blank, no precipitation, average wind speed is less than 8 m/s, without clouds;
(2) From the point of observation data, there is no obvious disturbance process.

This paper only gives the analysis of Spatial Electric Field in the altitude of 0 ~ 30 km.

The following empirical formula can give the law of Atmospheric electric field intensity and altitude [5]:

$$\begin{aligned} &E(z) = E(0)\exp(-az + bz^2) && (0 \sim 10km) \\ &E(z) = E(Z_0)\exp(-cz) && (10 \sim 30km) \end{aligned} \tag{1}$$

In the equations, z means altitude, E(z) is for the unit of V/m, the direction for the vertical is downward.

Parameters of Spatial Electric Field are a=0.591, b=0.0261, and c=0.124.

Among the parameters, the E (0) means that the main electric field intensity in the altitude of 0. $E(Z0)$ can be got by empirical formula. $E(0)$ in different places is different, so the method of observation to get all of E(0) doesn't work. The calculate algorithm of E (0) is described as below:

Suppose geographical coordinates (longitude ξ_i 、 latitude θ_i) ($i \in [1, N]$ and i is an integer) and their electric field strength value in 0 altitude $E(0)_i$ which is got by observation are known worldwide within any N (N>5) point (n_1 、 n_2 、 $n_3 \dots n_N$).

Using the method as follows to get any other electric field strength value in 0 altitude $E(0)_m$ at point m (longitude ξ_m 、 latitude θ_m):

Calculating the distance d_i (d_1 、 d_2 、 $d_3 \dots d_N$) from point m to each reference point (1..N);

Selecting the Minimum of five values (d_a 、 d_b 、 d_c 、 d_d 、 d_e) from d_i (d_1 、 d_2 、 d_3 … d_N) and their geographical coordinates (n_a 、 n_b 、 n_c 、 n_d 、 n_e);

Getting the electric field strength value in 0 altitude ($E(0)_a$ 、 $E(0)_b$ 、 $E(0)_c$ 、 $E(0)_d$ 、 $E(0)_e$)of (n_a 、 n_b 、 n_c 、 n_d 、 n_e) by searching the data of observation;

Using the method of weighted average to get the electric field strength value in 0 altitude of point m --- $E(0)_m$. The weights are $\left\{\frac{1}{d_a} \quad \frac{1}{d_b} \quad \frac{1}{d_c} \quad \frac{1}{d_d} \quad \frac{1}{d_e}\right\}^T$;

Calculating the estimating field strength value $E(0)_m$ in 0 altitude of point m:

$$E(0)_m = \begin{Bmatrix} E(0)_a \\ E(0)_b \\ E(0)_c \\ E(0)_d \\ E(0)_e \end{Bmatrix}^{\mathrm{T}} \cdot \begin{Bmatrix} e_a / d_{\mathrm{a}} \\ e_b / d_b \\ e_c / d_c \\ e_d / d_d \\ e_e / d_e \end{Bmatrix} \cdot \frac{1}{e_a / d + e_b / d_b + e_c / d_c + e_d / d_d + e_e / d_e} \qquad (2)$$

Here, $[e_a, e_b, e_c, e_d, e_e]^T$ is known as zero altitude correction coefficient of electric field. If the number of observation points in the model are too small, or the analyzed point is at the edge of two different regions characteristics electric field, $[e_a, e_b, e_c, e_d, e_e]^T$ should subject to the following condition: $e_a \geq e_b \geq e_c \geq e_d \geq e_e$.

It's easy to see that the more number of observed data (N value increases) is collected, the estimating value will be more accurate. If $E(0)_m$ is available, can according to the empirical formula calculate any point in the spatial of the electric field values.

3 The Model of Spatial Electric Field Changing over Time

3.1 The Model of Spatial Electric Field Changing over Season

The year is divided into summer half year (from May to October) and winter half year (November ~ April). The average electric field strength in 0 altitude of summer half year is lower than that of winter half year.

Suppose the average electric field strength in 0 altitude of summer half year is $E(0)_{i1}$. Suppose the average electric field strength in 0 altitude of winter half year is $E(0)_{i2}$. Suppose geographical coordinates (longitude ξ_i 、 Latitude θ_i) ($i \in [1, N]$ and i is an integer) and their electric field strength value in 0 altitude $E(0)_i$ which is got by observation are known worldwide within any N (N>5) point (n_1 、 n_2 、 $n_3 \ldots n_N$).

The method of calculating $E(0)_{i1}$ and $E(0)_{i2}$ is as follows:

The following relationship can be listed at below:

$$E(0)_{i2} = \lambda \cdot E(0)_{i1} \tag{3}$$

$$E(0)_{i2} + E(0)_{i1} = 2E(0)_{i} \tag{4}$$

λ is relation coefficient which is equal to 1.1. According to (1) and (2), $E(0)_{i1}$ and $E(0)_{i2}$ can be calculated as follows:

$$E(0)_{i1} = 0.952E(0)_{i} \tag{5}$$

$$E(0)_{i2} = 1.048E(0)_{i} \tag{6}$$

3.2 The Model of Spatial Electric Field Changing in a Day

According to the electric field data in sunny day, we can find that electric field in sunny day near the ground atmospheric changes basically presented double peak-valley type, the peaks appear at about 07:00 and 23:00, and the valleys appear at about 05:00 and 12:00. This is basically consistent with the result from a lot of observations on land.

The spatial electric field variation simulation curve can is given in Fig. 1.

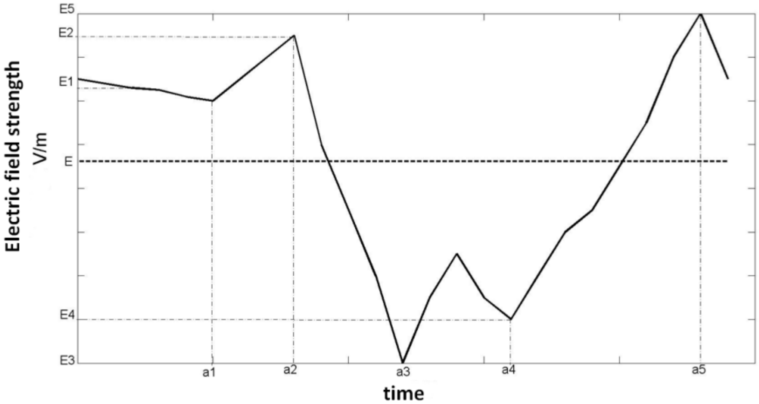


Fig. 1 The simulation curves of electric field on sunny days

As shown in Fig. 1, the dotted line represents daily average electric field data, the solid curve represents the changing of electric field data. The hour of 0 ~ 24 can be divided into five periods, the electric field changing law presents the five changes:

Step 1: 0~ a_1 (The period of stability): stable electric field intensity value is E_1;

Step 2: a_1 ~ a_2 (The period of a slight rise): in this period, electric field intensity is rising to the first peak;

Step 3: $a_2 \sim a_3$ (The period of a sharp fall): in area, in the valley fell to the first field intensity values;

Step 4: $a_3 \sim a_4$ (The period of oscillation) period: in this period, electric field intensity appears a slight oscillation, in the valley fell to the second electric field intensity values;

Step 5: $a_4 \sim a_5$ (The period of a sharp rise): in this period, electric field intensity grows up to the second peak.

If the daily average amplitude of electric field E1 is 1, the E1 ~ E5 is respectively sunny field amplitude coefficient, a1 ~ a5 is respectively the time coefficient of inflection point in field changes.

The method to get a1 ~ a5 and E1 ~ E5 are basically the same with the method of calculating average electric field per day in Section 2.1. The method can be solved under the condition of the following conditions by using interpolation:

$$\begin{cases} 0 < a_1 < a_2 < a_3 < a_4 < a_5 \\ 0 < E_3 < E_4 < E_1 < E_2 < E_5 \end{cases} \tag{7}$$

Nearest 5 points from the point being analyzed are selected, their time coefficient matrix and the amplitude coefficient matrix are known by observation as follows:

$$A = \begin{Bmatrix} a_{a1} & a_{a2} & a_{a3} & a_{a4} & a_{a5} \\ a_{b1} & a_{b2} & a_{b3} & a_{b4} & a_{b5} \\ a_{c1} & a_{c2} & a_{c3} & a_{c4} & a_{c5} \\ a_{d1} & a_{d2} & a_{d3} & a_{d4} & a_{d5} \\ a_{e1} & a_{e2} & a_{e3} & a_{e4} & a_{e5} \end{Bmatrix} \quad E = \begin{Bmatrix} E_{a1} & E_{a2} & E_{a3} & E_{a4} & E_{a5} \\ E_{b1} & E_{b2} & E_{b3} & E_{b4} & E_{b5} \\ E_{c1} & E_{c2} & E_{c3} & E_{c4} & E_{c5} \\ E_{d1} & E_{d2} & E_{d3} & E_{d4} & E_{d5} \\ E_{e1} & E_{e2} & E_{e3} & E_{e4} & E_{e5} \end{Bmatrix} \tag{8}$$

Among them, a_{ij} and E_{ij} are on behalf of serial number (a, b, c, d, e) of the observation point.

As for the vector $[a_1, a_2, a_3, a_4, a_5]^T$ and $[E_1, E_2, E_3, E_4, E_5]^T$, the method of calculating is as follows:

$$\begin{Bmatrix} a_1 & & & & \\ & a_2 & & & \\ & & a_3 & & \\ & & & a_4 & \\ & & & & a_5 \end{Bmatrix} = f(A_0 \cdot A)$$, function $f(X)$ makes Matrix X only to return diagonal elements and the other elements are zero.

$A_0 = \begin{Bmatrix} \alpha_{a1} & \alpha_{b1} & \alpha_{c1} & \alpha_{d1} & \alpha_{e1} \\ \alpha_{a2} & \alpha_{b2} & \alpha_{c2} & \alpha_{d2} & \alpha_{e2} \\ \alpha_{a3} & \alpha_{b3} & \alpha_{c3} & \alpha_{d3} & \alpha_{e3} \\ \alpha_{a4} & \alpha_{b4} & \alpha_{c4} & \alpha_{d4} & \alpha_{e4} \\ \alpha_{a5} & \alpha_{b5} & \alpha_{c5} & \alpha_{d5} & \alpha_{e5} \end{Bmatrix}$, A_0 is time factor transformation matrix. Each element of the matrix such as α_{ij} is the weight of time factor a_{ji}.

$$\begin{Bmatrix} E_1 & & & & \\ & E_2 & & & \\ & & E_3 & & \\ & & & E_4 & \\ & & & & E_5 \end{Bmatrix} = f(E_0 \cdot E)\text{'}\ E_0 = \begin{Bmatrix} \eta_{a1} & \eta_{b1} & \eta_{c1} & \eta_{d1} & \eta_{e1} \\ \eta_{a2} & \eta_{b2} & \eta_{c2} & \eta_{d2} & \eta_{e2} \\ \eta_{a3} & \eta_{b3} & \eta_{c3} & \eta_{d3} & \eta_{e3} \\ \eta_{a4} & \eta_{b4} & \eta_{c4} & \eta_{d4} & \eta_{e4} \\ \eta_{a5} & \eta_{b5} & \eta_{c5} & \eta_{d5} & \eta_{e5} \end{Bmatrix} \tag{9}$$

Where, E_0 is amplitude coefficient transformation matrix, Each element of the matrix such as η_{ij} is the weight of amplitude factor E_{ji}.

$$\begin{Bmatrix} \eta_{aj} \\ \eta_{bj} \\ \eta_{cj} \\ \eta_{dj} \\ \eta_{ej} \end{Bmatrix} = \begin{Bmatrix} \eta^*_{aj} / d_{\mathrm{a}} \\ \eta^*_{bj} / d_b \\ \eta^*_{cj} / d_c \\ \eta^*_{dj} / d_d \\ \eta^*_{ej} / d_e \end{Bmatrix} \cdot \frac{1}{\eta^*_{aj} / d_{\mathrm{a}} + \eta^*_{bj} / d_b + \eta^*_{cj} / d_c + \eta^*_{dj} / d_d + \eta^*_{ej} / d_e} \tag{10}$$

Where, j=1, 2, 3, 4, 5; $[\eta^*_{aj}, \eta^*_{bj}, \eta^*_{cj}, \eta^*_{dj}, \eta^*_{ej}]^T$ is called spatial correction coefficient。

If the total observation points of the model is too small, or the point being analyzed is just at the edge of two different regions of the electric field, we suggest that $\eta^*_{aj} \geq \eta^*_{bj} \geq \eta^*_{cj} \geq \eta^*_{dj} \geq \eta^*_{ej}$.

$$\begin{Bmatrix} \alpha_{aj} \\ \alpha_{bj} \\ \alpha_{cj} \\ \alpha_{dj} \\ \alpha_{ej} \end{Bmatrix} = \begin{Bmatrix} \alpha^*_{aj} / d_{\mathrm{a}} \\ \alpha^*_{bj} / d_b \\ \alpha^*_{cj} / d_c \\ \alpha^*_{dj} / d_d \\ \alpha^*_{ej} / d_e \end{Bmatrix} \cdot \frac{1}{\alpha^*_{aj} / d_{\mathrm{a}} + \alpha^*_{bj} / d_b + \alpha^*_{cj} / d_c + \alpha^*_{dj} / d_d + \alpha^*_{ej} / d_e} \tag{11}$$

Where, j=1, 2, 3, 4, 5; $[\alpha_{aj}^{*},\alpha_{bj}^{*},\alpha_{cj}^{*},\alpha_{dj}^{*},\alpha_{ej}^{*}]^{T}$ is called time correction coefficient。

If the total observation points of the model is too small, or the point being analyzed is just at the edge of two different regions of the electric field, we suggest that $\alpha_{aj}^{*} \geq \alpha_{bj}^{*} \geq \alpha_{cj}^{*} \geq \alpha_{dj}^{*} \geq \alpha_{ej}^{*}$.

4 Experiments and Experimental Result

4.1 The Formation of Thunderstorm

After the ground absorbed sun radiation, there is strong atmospheric convection. It tends to occur before the rainfall, and the lightning flash is divided into intra-cloud lightning flash and cloud-to-ground lightning flash. Both the intra-cloud lightning flash and the cloud-to-ground lightning flash happen in a vertical cylinder area, the distribution of the charge in this area is different from that in clear sky, which lead to thunderstorm [6].

Fig. 2 is the charge distribution diagram in the cylinder when thunderstorm happens.

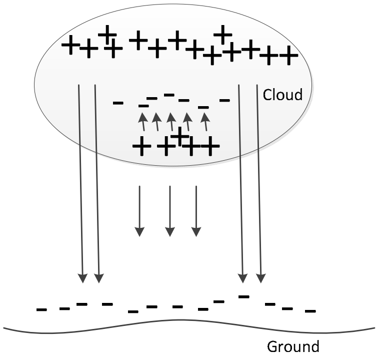


Fig. 2 the charge distribution diagram in the cylinder

As shown in Fig. 2, the distribution of the charge is in the cylinder is "positive-negative-positive-negative" from top to bottom. When the local electric field is about 400 kV/m , the lightning flash is produced. The intra-cloud lightning flash is generated by the negative charge at the medium of the cloud and the positive charge at the bottom of the cloud. And the cloud-to-ground lightning flash is generated by the positive charge at the bottom of the cloud and the negative charge of the earth.

4.2 The Influence of Cloud Flash

The influence factors of the lightning flash to the electric field are: thunderstorms, thunderstorm cloud height, the intensity of the thunderstorm, the duration of the thunderstorm, the charge distribution within the area of cylinder thunderstorms.

According to the Observation data, we found that in a horizontal direction the charge distribution is uncertainty, when the lightning flash happens, the spatial electric field widely changes in a concussion type, and the concussion cycle in about 20 minutes or so.

4.3 The Model of Spatial Electric Field in Thunderstorm

The influence of thunderstorms to the electric field vector spatial is not linear, even fuzzy and random, which causes great difficulty to modeling it.

A new modeling method is proposed: the electric field at each location (x, y, z) changes in a concussion type during thunderstorms occurred, and the ranges of x_m, y_m, z_m are as follows:

$$\begin{aligned} &x \in (x_m - x_{*1}, x_m + x_{*2}) \\ &y \in (y_m - y_{*1}, y_m + y_{*2}) \\ &z \in (z_m - z_{*1}, z_m + z_{*2}) \end{aligned} \tag{12}$$

In x_{**}, y_{**} and z_{**}, the first number of the Subscripts is the scale of the sudden environmental change, and 1 is serious, 2 is medium, 3 is mild. The second number is the range of the electric field when the environmental suddenly changes, and 1 is the lower limit, 2 is the upper limit. The influence of thunderstorm to the electric field at the z direction is strongest.

5 Conclusion

This paper proposes a spatial Electric field modeling method to generate the electric field data or database at a space and a certain time. Except for sunny day, the model at thunderstorm is provided, which is used for the virtual experiment. The influence of lightning flash to the spatial electric field is analyzed.

Acknowledgement The authors would like to thank Harbin Institute of Technology in China, for the financial support of this research.

References

1. Matsuo, T.: Electromagnetic Field Computation Using Space-Time Grid and Finite Integration Method. IEEE Transactions on Magnetics **46**(8), 3241–3244 (2010)
2. Zhou, H., Thottappillil, R., Diendorfer, G.: Calculation of electromagnetic fields in free space when lightning strikes a tall object. In: International Symposium on Lightning Protection, pp. 91–95 (2013)
3. Li, L., Tang, Y., Kou, B., et al.: Design and analysis of ironless linear electromagnetic launcher with high thrust density for space platform. In: International Symposium on Electromagnetic Launch Technology, pp. 1–6 (2012)
4. Wu, T., Lu, W.T., Liu, X.Y., et al.: Characteristics of Atmospheric Electric Field near the Earth's Surface Under Different Weather Conditions in Beijing. Journal of Applied Meteorological Science, 394–401 (2009)
5. Bin, X., Yin, Z., Huang-yu, J., et al.: Periodic Variations of Fair-Weather Atmospheric Electric Field Near Ground in Yangbajin of Tibet. Plateau Meteorology **28**(2), 314–318 (2009)
6. Fang, M., Tengfei. Z., Xin. W., et al.: Atmospheric electric field characteristics and CG lighting activities in the thunderstorm. Meteorological Hydrological & Marine Instruments, 117–121 (2011)

An Investigation on Maximum Entropy Estimation Based on Chrestenson Transform

Mingyong Zhou, Zhongkan Liu and Hiromitsu Hama

Abstract In this paper, we re-visit the topics of maximum entropy estimation in the sense of Chrestenson transform. As we all well know, maximum entropy estimation in the sense of Fourier's transform was well investigated and well-known Levinson-Burg algorithm was proposed and is widely used. We first elucidated the maximum entropy estimation based on [2], and then based on [2] derived and indicated the conditions where entropy rate h, as defined in the paper, converges and exists. Finally, we derive a general maximum entropy estimation for Chrestenson transform, and it shows that Levinson-Burg algorithm still applies in the sense of Chrestenson Transform. We also show that when the sampling data is p, an actual and complete spectral estimation can be obtained by a fast algorithm which is in line with part results in [2].

1 Introduction

Power spectral estimation is an important topics in signal processing. Maximum entropy estimation as one of important estimation method provides an approach to estimate signal power spectrum from finite observed samples and is widely used in practice in the sense of Fourier transform. In this paper, we re-investigate reference [2], and show several important practical observations based on maximum entropy estimation derivation in the sense of Chrestenson transform.

M. Zhou(✉)
Beihai College, Beihang University, Beijing, China
e-mail: zed6641@hotmail.com

Z. Liu
Beihang University, Beijing, China

H. Hama
Osaka City University, Osaka, Japan

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_8

2 Preliminary

Definition 1. (p-adic multiplication) for any integer p≥2, for any x∈[0, +∞), t∈[0, +∞),their p-adic integers can be represented as:

$$x=\left(x_{-s}\ldots x_{-2}\,x_{-1}\,x_0\bullet x_1\,x_2\ldots x_{r+1}\right)$$

$$t=\left(t_{-r}\ldots t_{-2}t_{-1}t_0\bullet t_1t_2\ldots t_{s+1}\right)$$

p-adic multiplication of t and x is defined as

$$t\otimes_p x=\sum_{k=-s}^{r+1}t_{1-k}x_k \tag{1}$$

where $\sum_{k=-s}^{r+1}$ denotes mod p sum.

Definition 2. (Chrestenson forward transform)

$$C(k,\omega)\underline{\underline{\Delta}}\exp\left\{-\frac{2\pi j}{p}\omega\otimes_p k\right\}\quad,\qquad k=0,1,2\cdots\infty \tag{2}$$

is defined as Chrestenson function. For a series $x(n),(n=0,1,\infty)$, its Chrestenson forward transform is defined as

$$X(\omega)\underline{\underline{\Delta}}\sum_{n=0}^{\infty}x(n)\exp\left\{-\frac{2\pi j}{p}\omega\otimes_p n\right\}\quad,\qquad \omega\in[0,1) \tag{3}$$

Where p is an integer number larger or equal to 2. Note that in [2] a prime number is required, and in fact in reference [1] it is shown that p can be any integers larger than 2. $\otimes_p$ denotes a p-adic multiplication defined in Definition 1.

Definition 3. (P-adic process) {x(n)} is defined as a p-adic stochastic process if auto-correlation $r(i)=E\left\lfloor x(k)\overline{x(l)}\right\rfloor$ depends only upon k -p l, where -p denotes mod p substraction.

For example, suppose $r(0)=a_0, r(1)=a_1+jb_1, r(2)=0, r(3)=0\cdots$, a simple auto-correlation matrix of p-adic stochastic process generated from p samples data is as follows:

$$A = \begin{bmatrix} a_0 & a_1 + jb_1 & 0 & \dots & 0 & a_1 - jb_1 \\ a_1 - jb_1 & a_0 & a_1 + jb_1 & \dots & 0 & 0 \\ \dots & \dots & & \dots & & \\ \dots & \dots & & \dots & a_0 & a_1 + jb_1 \\ a_1 + jb_1 & 0 & 0 & \dots & a_1 - jb_1 & a_0 \end{bmatrix} \tag{5}$$

Where A is a p x p matrix ,and $r(0) = a_0, r(1) = a_1 + jb_1, r(2) = 0, r(3) = 0 \cdots$.

Actually a practical method to construct the p-adic process is proposed in Section 5.

Definition 4.(Chrestenson power spectrum):

$$S(\omega) \underline{\underline{\Delta}} \sum_{\tau=0}^{\infty} r(\tau) \exp\left\{ -\frac{2\pi j}{p} \omega \otimes_p \tau \right\} \tag{6}$$

$$r(\tau) = \int_0^1 S(\omega) \exp\left\{ \frac{2\pi j}{p} \omega \otimes_p \tau \right\} d\omega, (\tau = 0,1,...p-1), \tag{7}$$

where $r(\tau)$ is auto-correlation defined in Definition 3.

3 Maximum Entroy Rate Computation

For p^m samples data , where m is any integer that is larger than 1, if it is p-adic process as defined in Definition 3, its auto-correlation matrix p^m x p^m can be easily shown to be of the following p^m x p^m form, where matrix A is defined in Eq.(5).

$$C = \begin{bmatrix} A, A, A..A \\ A, A, A...A \\ \dots \\ A, A, A...A \end{bmatrix}, \tag{8}$$

Note that from Definition 3, auto-correlation matrix is different from [2] Following the computation in [2], we can easily derive the following theorem.

Theorem 1. For matrix A in Eq.(5) and auto-correlation matrix C in Eq.(8), all the eigenvalues of auto-correlation matrix C is $m\lambda_1, m\lambda_2, ..., m\lambda_p$, where

$\lambda_1, \lambda_2, ..., \lambda_p$'s are the eigenvalues of matrix A, and $m\lambda_1, m\lambda_2, ..., m\lambda_p$'s eigenvalues are each repeated p^m .

The proof can be easily obtained by following the computations in [2].

4 Discussions on the Entropy Rate

We follow the definitions in [2], and define entropy rate:

Definition 5. (Entropy Rate)

$h = \lim H / p^m , (m- > \infty, p > 2)$

Theorem 2: From Definition 5 on entropy rate,

$$h = \ln(m^p) + \ln(\lambda_1 \lambda_2 ... \lambda_p), (m- > \infty, p > 2) \tag{9}$$

The proof can be obtained by direct computation.

From Theorem 2, we can observe 2 important points,

1, only when m=1, entropy rate h converges and exists.

2, when $m- > \infty$, regardless of p , entropy rate h does not converge.

Theorem 3. when m=1, given any integer p>2, the entropy rate is as follows

$$h = \frac{1}{p} \int_0^1 \ln[S(\omega)] d\omega$$

The proof is derived from theorem 2.

5 Construct P-adic Process by Extrapolation Method

For any integer p which is larger than 2, for convenience in this paper we assume $p = 2^m$, for example $p = 2^{10} = 1024$, for a real data x(1),x(2),…x(p/2), if the data are finite correlated, that is, r(l)=0, for certain l<p/2, it can be easily verified that the "image" extrapolation of data x(1),x(2),…x(p/2) by figure 1 method constructs a p-adic process , that is, the auto-correlation matrix after "image" extrapolation satisfies matrix A in Eq(5) .

[1]Figure 1 demonstrates a way how to "image" extrapolate the real data into an infinite data which construct a p-adic process.

---

[1] Extrapolation method is especially useful for applications where the data correlation needs to be removed. In maximum entropy estimation case, we only need m=1, that is one triangular data.

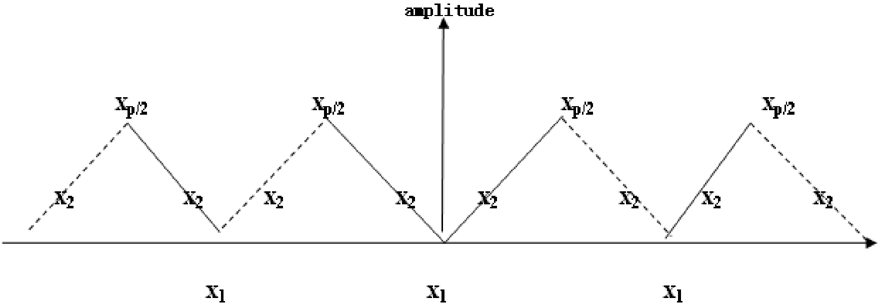


Fig. 1 "Image" extrapolation of data $x_1, x_2 \ldots x_{p/2}$ into infinite series.

6 Levinson-Burg Algorithm for Chrestenson Power Spectrum Estimation

From theorem 3, we can derive the maximum entropy estimation for Chrestenson transform for m=1. Following the argument as in Fourier's case, we can easily derive the maximum entropy estimation $S(\omega)$ as follows: given 2q+1 auto-correlation data r(0),r(1),r(-1)…r(q), r(-q), (2q+1<p), find $S(\omega)$ that makes the following objective function maximum.

$$J[S(\omega)] = \int_0^1 \ln[S(\omega)]d\omega + \sum_{k=-q}^{q} \alpha_k [r(k) - \int_0^1 S(\omega)\exp\left\{\frac{2\pi j}{p}\omega \otimes_p k\right\} d\omega] \quad (10)$$

The result can be derived as

$$S(\omega) = 1/(1 + \sum_{k=-q}^{q} \alpha_k \exp\left\{\frac{2\pi j}{p}\omega \otimes_p k\right\}) \quad (11)$$

Obviously $\alpha_k, (k = -q, ..0, ...q)$ can be derived from Levinson's algorithm given 2q+1 auto-correlation data r(0),r(1),r(-1)…r(q), r(-q), Or Levinson-Burg's algorithm can be applied for 2q+1 samples, where 2q+1<p.

Finally we can show that for p samples where p>2, a fast algorithm can be obtained for accurate Chrestenson power spectral estimation, which is definitely maximum entropy estimation as well.

$$S(\omega) \underset{=}{\Delta} \sum_{\tau=0}^{p-1} r(\tau) \exp\left\{-\frac{2\pi j}{p}\omega \otimes_p \tau\right\}, (p > 2) \quad (12)$$

When p is the exponential of 2, a fast algorithm can be easily proposed which is not included in this paper.

7 Conclusion

Maximum entropy estimation in the sense of Chrestenson transform is investigated in this paper. It is based on [2], and an entropy rate is derived which is slightly different from [2]. It is shown that only when m=1, regardless of p, the entropy rate under the sense of Chrestenson transform converges and exists. Finally for m=1 case, power spectrum estimation based on maximum entropy is investigated, and it is shown that Levinson-Burg algorithm still applies for Chrestenson spectral estimation when observed data is less than p, where p is any integer larger than 2.

References

1. Chrestenson, H.F.: A Class of gerneralized Walsh functions. Pacific Journal of Math. **5**, 17–31 (1955)
2. Qi, W., Zhongkan, L.: Computation of Chrestenson Maximum Entropy Spectral Estimation. Acta Electronics Sinica **21**(4), April 1993 (in Chinese)
3. Zhou, M., Liu, Z., Hama, H.: A Resolution-Controllable Harmonical Retrieval Approach on the Chrestenson Discrete Space. IEEE Transactions on Signal Processing **42**(5), May 1994

A Niching Genetic Algorithm Including an Inbreeding Mechanism for Multimodal Problems

Atsushi Ueno, Naoki Hagita and Tomohito Takubo

Abstract In this paper, we propose a new niching genetic algorithm (GA) for multimodal combinatorial optimization problems, which performs an effective search using a lot of small subpopulations. For this purpose, we combine two algorithms: the Adaptive Neighboring Search with Neighboring List 2 (ANS(List2)) by Nagata and the Parameter-free GA (PfGA) by Kizu *et al.* ANS(List2) performs self-distributive and locally adaptive search even for problems with complicated fitness landscapes, but it has a problem in which the numbers of individuals in some subpopulations become insufficient for effective searches. The PfGA has the ability to control the size and position of a small population adaptively. By incorporating the generation alternation model used in the PfGA into ANS(List2), our proposed algorithm can perform an effective search in a highly multimodal search space. We apply it to the Quadratic Assignment Problem and show its effectiveness.

Keywords Genetic algorithm · Niching method · Multimodal combinatorial optimization problem · Adaptive neighboring search · Parameter-free genetic algorithm

1 Introduction

Genetic Algorithms. (GAs) are biologically inspired search heuristics. They can find optimal or quasi-optimal solutions for problems that are difficult to be solved rigorously. Many of such problems have multimodal fitness landscapes, where there are many local solutions in the problem space. *Niching GAs* have been developed for difficult problems whose fitness landscapes are complicated and have deep narrow valleys. In a niching GA, a population of candidate solutions is divided into multiple subpopulations. Maintaining and dispersing the subpopulations is important for the success and effectiveness of search.

A. Ueno(✉) · N. Hagita · T. Takubo
Graduate School of Engineering, Osaka City University, 3-3-138, Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan
e-mail: ueno@info.eng.osaka-cu.ac.jp

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_9

Takahashi *et al.* [1] proposed a new algorithm called *Adaptive Neighboring Search* (ANS). In order to maintain the subpopulations, the ANS has introduced a new generation alternation model that is structured with a selection for reproduction by restricting mating individuals to neighbors and an elitist selection for survival within one centered parent and its children. The ANS realizes self-distributive and locally adaptive search even for problems with complicated fitness landscapes.

Nagata [2] proposed a new niching GA that is an improved algorithm from the ANS. The original algorithm in [1] handled optimization problems of real-valued functions. Nagata's algorithm has been revised for handling combinatorial optimization problems and introduced a new niche size equalization mechanism in order to increase the performance of maintaining and dispersing the subpopulations. This mechanism works based on neighboring individuals lists, each of which shows the neighboring individuals of each individual. The paper [2] proposed two definitions of neighborhood. One of them has a problem in which the numbers of individuals in some subpopulations become insufficient for effective searches.

In this paper, we propose a new niching GA to tackle this problem. We incorporate another generation alternation model used in the *Parameter-free GA* (PfGA) [3] into Nagata's algorithm. The PfGA has a variable-size population and needs at least only one individual to start or continue an effective search. By using this algorithm as the breeding mechanism in each subpopulation in Nagata's algorithm, we hope that effective searches are performed even in small subpopulations. By applying it to multimodal combinatorial optimization problems, we show that our algorithm can obtain good solutions with high probability.

2 Previous Works

2.1 Nagata's ANS Algorithm

In this subsection, we explain the ANS algorithm improved by Nagata [2]. The following pseudo code shows the algorithm, which is revised for handling combinatorial optimization problems from the original one in [1]:

```
procedure $ANS(N_p, N_c, K)$
begin
  Create the initial population, $\{x_1, x_2, ..., x_{N_p}\}$, randomly;
  repeat
    Create the neighboring list, $List_i$, for each $x_i$ $(i = 1, ..., N_p)$;
    Create a random permutation from 1 to $N_p$, $r[1], ..., r[N_p]$;
    for $s := 1$ to $N_p$ do
      $p_A := x_{r[s]}$;
      Select another parent, $p_B$, randomly from the neighboring list
      for $p_A$, $List_{r[s]}$;
      $\{c_1, c_2, ..., c_{N_c}\} := Breed(p_A, p_B)$;
      $x_{r[s]} := SelectBest(c_1, c_2, ..., c_{N_c}, x_{r[s]})$;
    end for
  until the end condition is satisfied;
  return the best individual in the population;
end.
```

This algorithm needs three parameters: N_p, N_c, and K. N_p is the population size, N_c is the number of children of each pair, and K is a parameter for deciding the neighborhood size. Procedure $Breed(p_A, p_B)$ in line 8 produces N_c children of p_A and p_B by performing crossover repeatedly. Procedure $SelectBest(c_1, c_2, ..., c_{N_c}, x_{r[s]})$ in line 9 returns one of the best individuals among p_A and its children. Only if the fitness of p_A is equal to that of p_B, these processes in line 8 and 9 are skipped and mutation takes place on $x_{r[s]}$. In this algorithm, each individual chooses its partner in its neighborhood and the survivor is selected also in its neighborhood. By this generation alternation model, maintenance and dispersion of the subpopulations are facilitated.

The paper [2] described three types of definitions of neighborhood: one had been proposed within the original ANS algorithm in [1] and the others were newly proposed. They are called, respectively, Neighboring List 1, 2, and 3.

The original one, Neighboring List 1, for each individual, x_i, is defined as follows:

$$List1_i = \{j \in \{1, ..., N_p\} \mid ord_{ij} < K\}, \tag{1}$$

where ord_{ij} is the ranking of each individual, x_j, in ascending order of distance to x_i, not including x_i itself. In the early stage of search, the sizes of neighborhoods are large because the individuals are well dispersed in the search space. As search progresses, the individuals are divided into dense subpopulations and the sizes of neighborhoods shrink self-adaptively. When using this definition, however, large subpopulations tend to pull surrounding individuals closer. Therefore, it is hard to maintain small subpopulations for a long term.

For this problem, Nagata proposed Neighboring List 2 and 3. Neighboring List 2 for each x_i is defined as follows:

$$List2_i = \{j \in \{1, ..., N_p\} \mid ord_{ij} < K, ord_{ji} < K\}. \tag{2}$$

This list is a subset of Neighboring List 1 for x_i. An element, j, is included in this list if x_j is in the neighborhood of x_i and x_i is also in the neighborhood of x_j. When using this definition, small subpopulations tend to be maintained steadily. Neighboring List 3 for each x_i is defined as follows:

$$List3_i = \{j \in \{1, ..., N_p\} \mid ord_{ji} < K\}. \tag{3}$$

An element, j, is included in this list if x_i is in the neighborhood of x_j. When using this definition, small subpopulations tend to pull surrounding individuals closer. Therefore, the sizes of the subpopulations tend to be equalized in the long run.

This algorithm was applied to the Job Shop Scheduling Problem (JSSP) and showed a good performance despite its simplicity. In comparing Neighboring List 1–3, the algorithm with Neighboring List 3 (ANS(List3)) has the best performance on the average. However, the algorithm with Neighboring List 2 (ANS(List2)) has close performance to it. ANS(List3) keeps a lot of subpopulations the sizes of which are about K. ANS(List2) keeps more subpopulations, although they include ones the sizes of which are considerably smaller than K. It seems that the decrease in performance of ANS(List2) is because the small subpopulations cannot search effectively even in their neighborhood.

2.2 PfGA

In this subsection, we explain the PfGA. The following pseudo code shows the algorithm:

```
procedure $PfGA()$
begin
1: Create the initial population, $S'$, including only one individual, which
   is selected randomly from the whole search space, $S$;
2: repeat
3:    if $|S'| = 1$ then add another individual selected randomly from $S$
      into $S'$;
4:    Pick parents, $P_1$ and $P_2$, randomly from $S'$ and produce children,
      $C_1$ and $C_2$, by crossover;
5:    Mutate one randomly chosen child, $C_1$ or $C_2$.
      $\{P_1, P_2, C_1, C_2\}$ is called "family";
6:    Select 1–3 individuals for survival among the family and place them
      back into $S'$;
7: until the end condition is satisfied;
8: return the best individual in the population;
end.
```

The process in line 6 is decided by the fitness values of the family members as follows:

Case 1: Both C_1 and C_2 are better than P_1 and P_2.
C_1, C_2, and the better parent are placed back into S', increasing $|S'|$ by one.
Case 2: Both P_1 and P_2 are better than C_1 and C_2.
Only the better parent is placed back into S', decreasing $|S'|$ by one.
Case 3: Only one of either P_1 or P_2 is better than C_1 and C_2.
The better parent and the better child are placed back into S', maintaining $|S'|$.
Case 4: Only one of either C_1 or C_2 is better than P_1 and P_2.
Only the better child is placed back into S' and one individual selected randomly from S is added into S', maintaining $|S'|$.

The paper [3] showed that the PfGA is a very compact and fast adaptive search algorithm accelerating its evolution based on a variable-size population and taking a balance between exploration and exploitation. This algorithm needs no preset parameters. It has only one small population and controls its size and position adaptively. It needs at least only one individual to start or continue an effective search.

3 A Niching GA Including an Inbreeding Mechanism

As mentioned in Sect. 2.1, ANS(List2) has a problem in which the numbers of individuals in some subpopulations become insufficient for effective searches. In this section, we will propose a new niching GA to tackle this problem. We incorporate the generation alternation model used in the PfGA into ANS(List2). We expect that effective searches are performed even in small subpopulations because the PfGA can control the size and position of a small population adaptively.

3.1 Basic Algorithm

We adopt the ANS algorithm explained in Sect. 2.1, procedure $ANS(N_p, N_c, K)$, as the framework of our algorithm. The generation alternation model used in the PfGA is implemented as two procedures in $ANS(N_p, N_c, K)$: $Breed(p_A, p_B)$ and $SelectBest(c_1, ..., c_{N_c}, x_{r[s]})$. In $Breed(p_A, p_B)$, the loop block in line 2–7 in procedure $PfGA()$ is executed. The process in line 6 in $PfGA()$ is simplified as follows:

1. The better parent is placed back into S' and the worse parent is eliminated.
2. Each of the children is placed back into S' if and only if it is better than the worse parent.

In Case 1 and 2 in Sect. 2.2, this simplified process works in the completely same manner, although it works a little differently in Case 3 (sometimes) and 4 (at all times). The end condition in line 7 in $PfGA()$ is that the number of children produced in the loop reaches N_c. It means that the number of iterations is $N_c/2$. After exiting the loop, one of the best individuals in S' other than p_B is returned as the return value of $SelectBest(c_1, ..., c_{N_c}, x_{r[s]})$.

3.2 Improvements for Maintaining Subpopulations

As mentioned in Sect. 2.1, ANS(List2) tends to maintain small subpopulations steadily. We introduce two additional improvements for maintaining subpopulations.

First, we improve the way of adding a new individual in line 3 in $PfGA()$. In the original algorithm in [3], the additional individual is selected randomly from the whole search space, S. In our proposed algorithm, it is selected randomly between p_A, p_B, and mutated individuals from them. The reason for this is that adding an individual distant from other members of a subpopulation causes it to be scattered and makes it difficult to maintain it.

Secondly, we improve the definition of the fitness function of individuals. In our algorithm, the fitness of each individual is related to the distance from p_A. Individuals close to p_A are given high fitnesses. Let us now consider a case when the original fitness function f of an individual q is always less than or equal to zero: $f(q) \leq 0$ $(\forall q \in S)$. This type of fitness function is called *cost function*. In this case, our algorithm uses the following fitness function:

$$g(q) = \begin{cases} \frac{f(q)-f(p_A)}{d(q,p_A)} & (f(q) > f(p_A),\ q \neq p_B), \\ \epsilon & (q = p_A), \\ -\epsilon & (q = p_B), \\ 0 & (otherwise), \end{cases} \tag{4}$$

where $d(q, p_A)$ is the distance between q and p_A, and ϵ is a small positive value[1]. By using this fitness function, individuals close to p_A have an advantage in surviving. This tendency is useful for keeping the search area near p_A.

These two improvements enhance the local search ability of the PfGA. This generation alternation model can be seen as a sort of "inbreeding mechanism" because it repeats mating between two individuals in a family of p_A, p_B, and their offsprings.

3.3 Improvements for Accelerating Exploration

As mentioned in Sect. 2.1, when the fitness of p_A is equal to that of p_B, Nagata's ANS algorithm performs mutation on p_A and skips breeding. In the same cases, our algorithm replaces p_A with an individual selected randomly from the whole search space and skips breeding. It means that it starts searching another new niche. Occurrence of the case where the two selected individuals have the same fitnesses indicates that the local search in their neighborhood is near convergence. Because our algorithm can perform local search with a small population, we expect that this redeployment mechanism will accelerate the whole search.

4 Experiments and Results

In this section, we apply our proposed algorithm to the *Quadratic Assignment Problem* (QAP) and verify its effectiveness.

4.1 QAP

The QAP is an NP-hard combinatorial optimization problem, which has a search space with a highly multimodal landscape. Given a set of N facilities, P, and a set of N locations, L, the problem is to find an assignment (bijection), $\phi : P \to L$, such that the total shipment cost is minimized. The total shipment cost of ϕ is given as follows:

$$C(\phi) = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} b_{\phi(i)\phi(j)} \quad (a_{ij} \geq 0,\ b_{\phi(i)\phi(j)} \geq 0), \tag{5}$$

---

[1] Because only p_B has a negative fitness, it can never survive. However, the individual selected as p_B is selected as p_A once in the for loop in procedure ANS. A local search in its neighborhood is performed at that time.

where a_{ij} is the flow (e.g. the amount of supplies transported) from facility i to facility j, and $b_{\phi(i)\phi(j)}$ is the distance between locations assigned to facilities i and j. $C(\phi)$ is greater than or equal to zero for any assignment ϕ.

We pick up all problem instances whose sizes are between 20 and 40 ($20 \leq N \leq 40$) from the QAP Library (QAPLIB) [4] and perform experiments on them.

4.2 Experimental Setting

We compare the performance of our algorithm with ANS(List2) and ANS(List3) in [2] and the original PfGA in [3]. In all algorithms, each chromosome is a permutation of the N locations, which represents an assignment ϕ. If the ith element in it is j, it means that facility i is assigned to location j. We use the order crossover (OX) and the 2-change mutation, which swaps two randomly chosen genes in a chromosome. In ANS(List2), ANS(List3), and the PfGA, the fitness is defined as the negative of the above total shipment cost: $-C(\phi)$. Because this value is always less than or equal to zero, we use it as $f(q)$ in (4) in our algorithm. ANS(List2), ANS(List3), and our algorithm need the definition of the distance between two individuals. We use the Hamming distance in all the algorithms. The parameters in our algorithm are set as follows: $N_p = 300$, $N_c = 100$, and $K = 5$ and 10. The parameters in ANS(List2) and ANS(List3) were tuned in preliminary experiments independently. In ANS(List2), $N_p = 600$, $N_c = 60$, and $K = 5$. In ANS(List3), $N_p = 1000$, $N_c = 60$, and $K = 5$. The PfGA needs no parameters.

For each problem instances and each algorithm, we perform 10 trials with different random seeds. All trials have a limit to computation time for the sake of fair comparison between the four algorithms. The limit is three minutes for small-size problems ($N < 30$) and six minutes for midsize problems ($N \geq 30$). The experiments are performed on the Intel®Core™2 Duo E8400 at 3.00GHz with 3.00GB RAM. All the algorithms are implemented by Java.

4.3 Results

Table 1 shows the results of the comparison experiments. The most left column, Instance (Opt), shows the names of the problem instances with the optimal or best-known solutions in parentheses. The other columns show the averages of the smallest costs, $C(\phi)$, found in each trial over 10 trials, with the number of trials in which the optimal or best-known solution was found in parentheses. The bold and shaded items indicate the best performance among the five settings. The underlined items indicate that the algorithm found the optimal of best-known solution in all 10 trials.

As shown in Table 1, our algorithm outperforms the previous algorithms. Its performances are better than (or equal to) those of the others in all the problem instances. In the case of $K = 5$, it found the optimal or best-known solution once at least in 52 problem instances out of 57 instances. It was found in all 10 trials in 37 problem instances. In the case of $K = 10$, it was found once at least in 51 problem instances and in all 10 trials in 39 problem instances.

Table 1 Results of the comparison experiments.

Instance (Opt)	Proposed GA (K = 5)	Proposed GA (K = 10)	ANS(List2)	ANS(List3)	PfGA
bur26a (5426670)	**5426670 (10)**	**5426670 (10)**	5427928.7 (7)	5428537.3 (4)	5438760 (1)
bur26b (3817852)	**3817852 (10)**	**3817852 (10)**	3817892.6 (6)	3818042.2 (6)	3829962.8 (0)
bur26c (5426795)	**5426795 (10)**	**5426795 (10)**	5426812.8 (6)	5426926.9 (3)	5439521.7 (0)
bur26d (3821225)	**3821225 (10)**	**3821225 (10)**	3821238.4 (5)	3821268 (3)	3823979.2 (0)
bur26e (5386879)	**5386879 (10)**	**5386879 (10)**	5386926.8 (8)	5386925.7 (8)	5393684.5 (0)
bur26f (3782044)	**3782044 (10)**	**3782044 (10)**	3782048.8 (8)	3782077.1 (8)	3791057.1 (0)
bur26g (10117172)	**10117172 (10)**	**10117172 (10)**	10117205 (6)	10117439 (4)	10139582 (1)
bur26h (7098658)	**7098658 (10)**	**7098658 (10)**	**7098658 (10)**	7098682.7 (9)	7112174.8 (3)
chr20a (2192)	**2194 (5)**	2202.8 (3)	2375 (0)	2384 (0)	2769.2 (0)
chr20b (2298)	**2327.6 (5)**	2337.4 (3)	2458.4 (0)	2441 (0)	2556.8 (0)
chr20c (14142)	**14142 (10)**	**14142 (10)**	14428.8 (7)	14343.8 (6)	22398.8 (0)
chr22a (6156)	6176.6 (6)	**6176.4 (7)**	6334.4 (0)	6345.2 (0)	6671.6 (0)
chr22b (6194)	6233 (4)	**6206.6 (8)**	6354 (0)	6412.8 (0)	6442.4 (0)
chr25a (3796)	3958.4 (2)	**3919.6 (2)**	4446.4 (0)	4528 (0)	5868.6 (0)
esc32a (130)	130.2 (9)	**130 (10)**	133 (2)	135.6 (0)	141.8 (0)
esc32b (168)	**168 (10)**	**168 (10)**	**168 (10)**	**168 (10)**	189.2 (0)
esc32c (642)	**642 (10)**	**642 (10)**	**642 (10)**	**642 (10)**	**642 (10)**
esc32d (200)	**200 (10)**	**200 (10)**	**200 (10)**	**200 (10)**	206.4 (3)
esc32e (2)	**2 (10)**	**2 (10)**	**2 (10)**	**2 (10)**	**2 (10)**
esc32f (2)	**2 (10)**	**2 (10)**	**2 (10)**	**2 (10)**	**2 (10)**
esc32g (6)	**6 (10)**	**6 (10)**	**6 (10)**	**6 (10)**	**6 (10)**
esc32h (438)	**438 (10)**	**438 (10)**	**438 (10)**	**438 (10)**	444.8 (0)
had20 (6922)	**6922 (10)**	**6922 (10)**	**6922 (10)**	**6922 (10)**	6945.6 (5)
kra30a (88900)	**88900 (10)**	**88900 (10)**	90323 (1)	90072 (2)	94410 (0)
kra30b (91420)	91441 (7)	**91420 (10)**	91942 (0)	91857 (2)	94861 (0)
kra32 (88700)	**88700 (10)**	**88700 (10)**	89872 (3)	90106 (1)	92591 (0)
lipa20a (3683)	**3683 (10)**	**3683 (10)**	3688.1 (9)	**3683 (10)**	3758.7 (1)
lipa20b (27076)	**27076 (10)**	**27076 (10)**	**27076 (10)**	**27076 (10)**	30182.9 (2)
lipa30a (13178)	**13178 (10)**	**13178 (10)**	13333 (2)	13292.8 (3)	13404.3 (0)
lipa30b (151426)	**151426 (10)**	**151426 (10)**	**151426 (10)**	**151426 (10)**	175540.8 (0)
lipa40a (31538)	31753.8 (3)	**31748.6 (3)**	31914.5 (0)	31922.1 (0)	31949.8 (0)
lipa40b (476581)	**476581 (10)**	**476581 (10)**	491319.1 (8)	484730.4 (9)	553552.5 (1)
nug20 (2570)	**2570 (10)**	**2570 (10)**	2570.8 (8)	2575.4 (5)	2623.8 (0)
nug21 (2438)	**2438 (10)**	**2438 (10)**	2440 (7)	2440 (6)	2499.2 (0)
nug22 (3596)	**3596 (10)**	**3596 (10)**	**3596 (10)**	3597.2 (8)	3717.4 (0)
nug24 (3488)	**3488 (10)**	**3488 (10)**	3491.4 (8)	3497.4 (3)	3577 (0)
nug25 (3744)	**3744 (10)**	**3744 (10)**	3746.6 (4)	3748.2 (2)	3832.4 (0)
nug27 (5234)	**5234 (10)**	**5234 (10)**	5265.4 (1)	5256.4 (3)	5357.6 (0)
nug28 (5166)	5167 (9)	**5166 (10)**	5210.4 (0)	5213.2 (0)	5300.8 (0)
nug30 (6124)	**6124.4 (9)**	6125.6 (6)	6146 (1)	6149.6 (1)	6323.6 (0)
rou20 (725522)	**725729.4 (4)**	725735.4 (3)	729242.8 (0)	730432.4 (0)	740563.8 (0)
scr20 (110030)	**110030 (10)**	**110030 (10)**	110139.6 (6)	110132.4 (6)	114279.6 (0)
ste36a (9526)	**9544 (6)**	9549.2 (4)	9721.6 (0)	9718.6 (0)	10289.2 (0)
ste36b (15852)	**15852 (10)**	**15852 (10)**	16180.6 (1)	16007.2 (2)	18343.6 (1)
ste36c (8239110)	8241746 (8)	**8241278.4 (8)**	8346629.2 (0)	8319322 (0)	8685196 (0)
tai20a (703482)	**705659.6 (2)**	706126.8 (0)	712408 (0)	711994.8 (0)	726504.4 (0)
tai20b (122455319)	**122455319 (10)**	**122455319 (10)**	**122455319 (10)**	**122455319 (10)**	123014928 (4)
tai25a (1167256)	**1174263.6 (0)**	1175441.6 (0)	1193423.8 (0)	1191365.6 (0)	1207531.8 (0)
tai25b (344355646)	**344355646 (10)**	**344355646 (10)**	344868846 (0)	344786166 (2)	366664265 (2)
tai30a (1818146)	1830362.4 (0)	**1830010.6 (0)**	1863023.8 (0)	1857918.8 (0)	1884255.8 (0)
tai30b (637117113)	**637117113 (10)**	**637117113 (10)**	638344151 (0)	637983674 (2)	685997236 (0)
tai35a (2422002)	2446436.4 (0)	**2444963.4 (0)**	2482488.2 (0)	2489810 (0)	2506985 (0)
tai35b (283315445)	283584237 (3)	**283404349 (7)**	284652824 (0)	284465007 (0)	297323967 (0)
tai40a (3139370)	3182996.8 (0)	**3180278 (0)**	3226448.4 (0)	3232603.2 (0)	3248956 (0)
tai40b (637250948)	**637250948 (10)**	**637250948 (10)**	641353534 (0)	641201394 (0)	683963110 (0)
tho30 (149936)	**149936 (10)**	149970.4 (9)	150710.8 (0)	150766.6 (0)	156524.8 (0)
tho40 (240516)	241262.8 (0)	**241138.8 (0)**	243576.6 (0)	243606.6 (0)	248701.2 (0)
Average	(7.93)	**(7.95)**	(4.28)	(4.00)	(1.12)

The PfGA has the worst performance in all the problem instances. We think that this is because the PfGA is not suitable for solving search problems with highly multimodal landscapes. In comparison between ANS(List2) and ANS(List3), the former has better performance in more cases than the latter, unlike the case in which they are applied to the JSSP shown in [2]. As mentioned in Sect. 2.1, ANS(List2) keeps more subpopulations, including small ones, than ANS(List3). If there are a lot of niches that can be searched by a small subpopulation, ANS(List2) has an advantage in global search because it keeps more subpopulations. We think that the QAP may have such landscapes.

In comparison between the cases of $K = 5$ and 10 in our algorithm, the latter has a little better performance averagely. The difference is, however, very small. The value $K = 10$ may be too large and some individuals in some subpopulations may become redundant on the grounds that both the values of K in the best settings of ANS(List2) and ANS(List3) are set to 5. As mentioned in Sect. 3.3, our algorithm has a mechanism for redeploying redundant individuals for searching new niches. We think that this mechanism is the main reason why the setting with $K = 10$ can perform as good as the setting with $K = 5$. Our algorithm can perform effective local searches by using the generation alternation model of the PfGA and an effective global search by using this redeployment mechanism.

5 Conclusions

In this paper, we have proposed a new niching GA for multimodal combinatorial optimization problems, which performs an effective search using a lot of small subpopulations. For this purpose, we have combined two algorithms: ANS(List2) and the PfGA. ANS(List2) performs self-distributive and locally adaptive search even for problems with complicated fitness landscapes, but it has a problem in which the numbers of individuals in some subpopulations become insufficient for effective searches. The PfGA has the ability to control the size and position of a small population adaptively. By incorporating the generation alternation model used in the PfGA into ANS(List2), our proposed algorithm can perform an effective search in a highly multimodal search space. We have applied it to the QAP instances and shown its effectiveness over ANS(List2), ANS(List3), and the original PfGA.

We are now trying to compare it with other state-of-the-art algorithms on the QAP and other multimodal combinatorial optimization problems such as JSSP.

References

1. Takahashi, O., Kimura, S., Kobayashi, S.: An Adaptive Neighboring Search using Crossover-like Mutation for Deceptive Multimodal Function Optimization. Trans. Japanese Society for Artificial Intelligence **16**(2), 175–184 (2001). (in Japanese)

2. Nagata, Y.: An Adaptive Niching Genetic Algorithm using a Niche Size Equalization Mechanism. Trans. Japanese Society for Artificial Intelligence **24**(1), 92–103 (2009). (in Japanese)
3. Kizu, S., Sawai, H., Adachi, S.: Parameter-free Genetic Algorithm (PfGA) using Adaptive Search with Variable-Size Local Population and Its Extension to Parallel Distributed Processing. IEICE Trans. Information and Systems **J82–D2**(3), 512–521 (1999). (in Japanese)
4. Burkard, R.E., Karisch, S.E., Rendl, F.: QAPLIB - A Quadratic Assignment Problem Library. http://www.opt.math.tu-graz.ac.at/qaplib/

Part II
Intelligent Data Security and Wearable Computing

Robust Reversible Colour Image Digital Watermarking Using Offset Coefficient Technique

Ching-Yu Yang and Wen-Fong Wang

Abstract By using offset coefficient of a host block, we design a robust reversible digital watermarking for color images. Based on pixel-value-offset of a host block, a watermark can be effectively embedded in a host image without noticeable visual distortion. Experiments indicate that the perceived quality generated by the proposed method is good while the payload size is superior to existing techniques. Moreover, the marked images generated by our method are resisting from manipulations such as brightness, contrast, color quantization, equalized, (edge) sharpening, inversion, JPEG, JPEG2000, noise additions, pixel-truncation, and winding.

Keywords Data hiding · Color image digital watermarking · Offset coefficient

1 Introduction

Recently, data hiding techniques are gradually attracted people and being used to secure private data and protect intellectual property right, content authentication, and copyright protection [1,2]. Generally, data hiding can be divided into two categories: steganography and digital watermarking. Conventional steganographic methods often provide a high payload for hiding secret message while maintaining

---

C.-Y. Yang(✉)
Department of Computer Science and Information Engineering,
National Penghu University of Science and Technology, Penghu, Taiwan
e-mail: chingyu@npu.edu.tw

W.-F. Wang
Department of Computer Science and Information Engineering,
National Yunlin University of Science and Technology, Yunlin, Taiwan
e-mail: wwf@yuntech.edu.tw

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_10

a good (or acceptable) perceived quality. The key feature of digital watermarking [3-4] is its robustness performance. Most of the watermarking schemes are capable of resisting from manipulations, such as compression, cropping, noise-addition, and quantization attacks. Since some host media such as medical images, geographic systems and satellite resources are priceless (or valuable), it is undesirable that the host media be damaged after data extraction. Several authors have presented robust lossless watermarking techniques [5-7] to achieve the goal. A major merit of a robust reversible data hiding technique is that in normal a hidden watermark can be extracted losslessly from a marked image while the host image can be restored completely. If the marked images were attacked by the third parties (or hackers), recovery of the host images would be failed, however, the extracted watermarks can still be recognized.

Ni et al. [5] proposed a robust reversible data hiding method based on patchwork theory, distribution features of pixel groups, error codes, and the permutation scheme. The payload size of this technique was limited by 1,024 bits, but the resultant images contained no salt-and-pepper noise and the average peak signal-to noise ratio (PSNR) exceeded 38 dB. Zeng et al. [6] suggested a lossless robust data hiding technique for color images. They embedded data bits into the host blocks by shifting mathematical difference values. Their experiments indicated that the resulting payload was 16 kb with the average PSNR was around 38 dB. Based on radius-weighted mean (RWM) norm, Yang and Wang [7] proposed a robust and reversible data hiding method for color images. Simulations confirmed that the resultant perceived quality is good while hiding capacity is not bad. The marked images generated by their method were tolerant of a variety of attacks.

In this paper, to further promote hiding storage, the authors propose a robust reversible data hiding method for color images based on offset coefficient of the blocks. The rest of the paper is organized as follows. The proposed bit-embedding and bit-extraction techniques are described in Section 2. Simulations are demonstrated in Section 3. Finally, a conclusion is summarized in Section 4.

2 The Proposed Method

Let H_k be the k-th block of size n × n taken from a host image, where $H_k = \Omega_x \cup \Omega_y$ with $\Omega_x = (\vec{\mathrm{X}}_{1j}, \vec{\mathrm{X}}_{2j})$ and $\Omega_y = (\vec{\mathrm{Y}}_{1j}, \vec{\mathrm{Y}}_{2j})$, as shown in Fig. 1, with n = 2. Also, let $\|\Omega_x\| = \left\| \left\lfloor (\vec{\mathrm{X}}_{1j} + \vec{\mathrm{X}}_{2j}) / 2 \right\rfloor \right\|$ and $\|\Omega_y\| = \left\| \left\lfloor (\vec{\mathrm{Y}}_{1j} + \vec{\mathrm{Y}}_{2j}) / 2 \right\rfloor \right\|$ be the Euclidean norm of Ω_x and Ω_y, respectively. The details of the proposed procedure for bit embedding and bit extraction are described in the following sections.

$\vec{X}_{1j}$	$\vec{Y}_{1j}$
$\vec{Y}_{2j}$	$\vec{X}_{2j}$

Fig. 1 A 2 × 2 host RGB block

2.1 Bit-Embedding

The main procedure for the proposed bit embedding is described in the following algorithm.

Algorithm 1. Hiding watermark bits in an RGB color image.

```
Input: An RGB host image $S=\{(r_i, g_i, b_i) \mid i=1,2,\ldots,|S|\}$, a scrambled
    watermark M, and three control parameters $\tau_r$, $\tau_g$, and $\tau_b$.
Output: A marked image $S'$ (and overhead bits).
Method:
Step 1. Input a block $H_k$, which derived from $S$. If the end of input is
    encountered, then proceed to Step 6.
Step 2. Compute $\|\Omega_x\|$ and $\|\Omega_y\|$ of $H_k$.
Step 3. Input a data bit $b_l$ from M.

Step 4. If $b_l = 0$, then do the following sub-steps:
    Step 4.1. If $\|\Omega_x\| \geq \|\Omega_y\|$ is satisfied, then do nothing, which means the block
        "carries" data bit 0 without altering the pixels' value, and proceed to Step
        1.
    Step 4.2. Otherwise, call the procedure "HidingCase0" and go back to Step 1.
Step 5. If $b_l = 1$, then do the following sub-steps:
    Step 5.1. If $\|\Omega_x\| < \|\Omega_y\|$ is satisfied, then do nothing, which means the block
        "carries" data bit 1 without altering the pixels' value, and proceed to Step
        1.
    Step 5.2. Otherwise, call the procedure "HidingCase1" and go back to Step 1.
Step 6. Stop.
```

The procedures of “HidingCase0” and “HidingCase1” are depicted in Figs. 2(a)-(b). Since overflow and underflow issues may arise during coefficient adjustment. To avoid underflow while the value of a target pixel (r_i, g_i, b_i) being reduced, the decrement is only conducted to r_i if $r_i \geq \tau_r$, to g_i if $g_i \geq \tau_g$, and to b_i if $b_i \geq \tau_b$. Similarly, to avoid overflow issue, the increment is only performed to r_i if $r_i \leq (255 - \tau_r)$, g_i if $g_i \leq (255 - \tau_g)$, and b_i if $b_i \leq (255 - \tau_b)$.

2.2 Bit-Extraction

The procedure of bit extraction is much simpler than that of bit embedding. Let $H'_k = \Omega'_x \cup \Omega'_y$ be the *j*-th block of size *n* × *n* taken from a marked image. Also let $\|\Omega'_x\|$ and $\|\Omega'_y\|$ be the Euclidean norm of Ω'_x and Ω'_y. The major steps of bit extraction are specified in the following algorithm.

Input: A marked image $S' = \{(r'_i, g'_i, b'_i) \mid i = 1, 2, \ldots, |S'|\}$ (and overhead information).

Output: A watermark *M* and restored host image *S*.

Method:

Step 1. Input a hidden block H'_k from S'. If the end of the input is encountered, proceed to Step 4.

Step 2. If H'_k is a skipped block, then go back to Step 1.

Step 3. Compute $\|\Omega'_x\|$ and $\|\Omega'_y\|$ of H'_k.

Step 3. If $\|\Omega_x\| \geq \|\Omega_y\|$ is satisfied, then data bit 0 is extracted, otherwise, data bit 1 is extracted, and go from Step 1.

Step 4. Assemble all the extracted bits, descramble the data bits, and form the watermark *M*, as well as restore host image *S*.

Step 5. Stop.

To help the decoder to successfully extract the hidden watermark and completely restore the host image, the overhead information, which used to record the position of the skipped blocks, has to be sent to the receiver. If the number of skipped blocks is N_s, which was determined below a threshold, say 2048, and the location of a skipped block is represented by $\left(\log_2\left\lfloor\frac{M}{n}\right\rfloor + \log_2\left\lfloor\frac{N}{n}\right\rfloor\right)$, then the overhead information can be computed by $\left(\log_2\left\lfloor\frac{M}{n}\right\rfloor + \log_2\left\lfloor\frac{N}{n}\right\rfloor\right) \times N_s \leq N_s \log_2 \frac{MN}{n^2}$ bits in total.

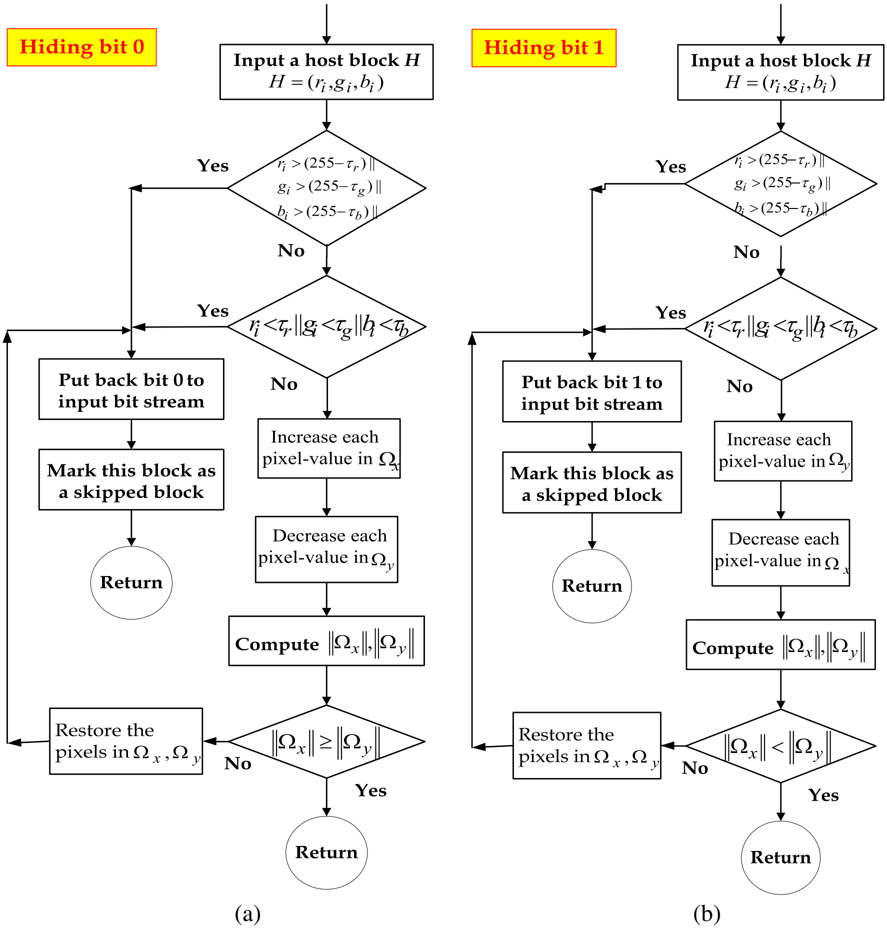

Hiding bit 0
Input a host block H
$H=(r_i,g_i,b_i)$
$r_i>(255-\tau_r)\|$
$g_i>(255-\tau_g)\|$
$b_i>(255-\tau_b)\|$
Yes
No
$r_i<\tau_r\|g_i<\tau_g\|b_i<\tau_b$
Put back bit 0 to input bit stream
Mark this block as a skipped block
Return
Increase each pixel-value in Ω_x
Decrease each pixel-value in Ω_y
Compute $\|\Omega_x\|,\|\Omega_y\|$
$\|\Omega_x\|\geq\|\Omega_y\|$
Restore the pixels in Ω_x,Ω_y
(a)
Hiding bit 1
Put back bit 1 to input bit stream
Increase each pixel-value in Ω_y
Decrease each pixel-value in Ω_x
$\|\Omega_x\|<\|\Omega_y\|$
(b)

Fig. 2 Flowcharts of the pixel-adjustment procedure. (a) HidingCase0 and (b) HidingCase1

3 Experimental Results

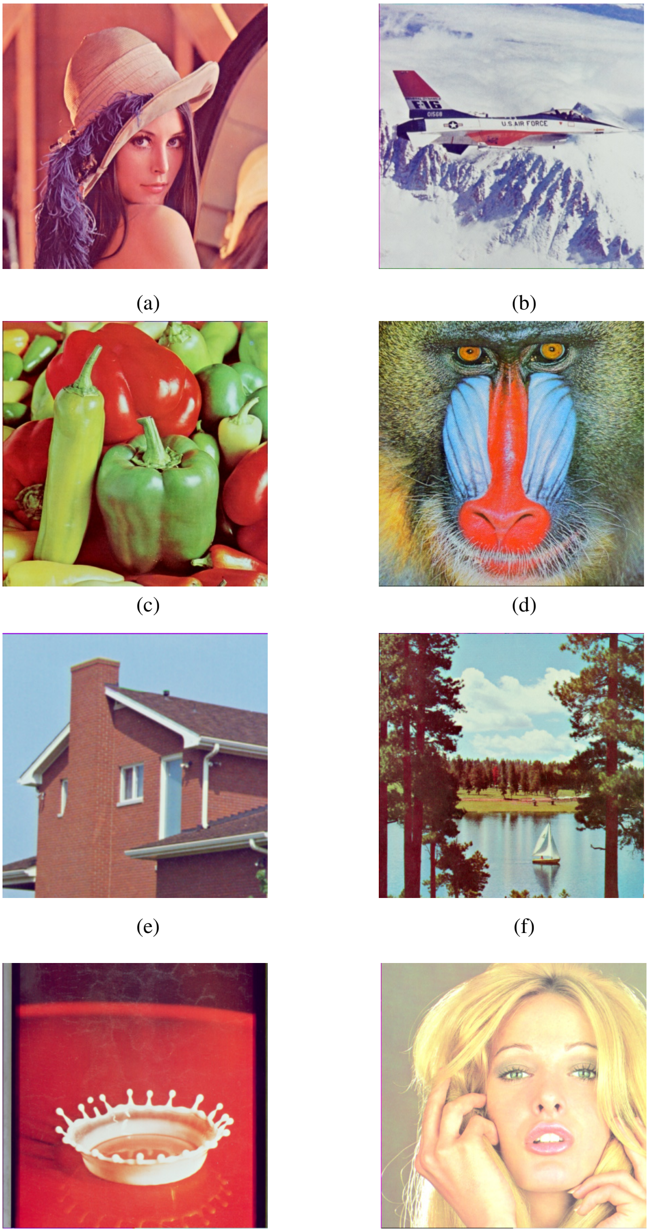
Several 512 × 512 color images were used as host images. Each RGB pixel of the host images is represented by 24 bits, 8 bits per component. The block size is 2 × 2. The marked images generated by the proposed method are depicted in Fig. 3. Notice that each host image equipped with a variety of control parameters ($\tau_r/\tau_g/\tau_b$) in the RGB system. For example, a set of control parameters, $\tau_r = 8$, $\tau_g = 5$, and $\tau_b = 8$ was used to introduce the marked image *Lena*. From the figure we can see that no noticeable color distortion appeared in the figures. Their corresponding PSNR/payload and the number of skipped blocks were given in Table 1. Although the average PSNR is moderate, the perceived quality is good. In addition, the average ratio of skipped and host blocks is below 1.32% (or 759 skipped block.) The PSNR is defined by $PSNR = 10 \times \log_{10} \frac{255^2}{MSE}$ with $MSE = (\sum_{i=1}^{MN} [(r_i - \hat{r}_i)^2 + (g_i - \hat{g}_i)^2 + (b_i - \hat{b}_i)^2]) / 3MN$. Here (r_i, g_i, b_i) and $(\hat{r}_i, \hat{g}_i, \hat{b}_i)$ denote the RGB pixel values of the host image and the marked image. Tradeoff between PSNR and payload for our method in four test image were shown in Fig. 4. The figure indicated that the image *Splash* provided the largest payload, while the image *Sailboat* introduced the best PSNR.

Performance comparison between our proposed method and existing schemes is given in Table 2. It can be seen that the proposed method provides the largest payload among these compared methods while the average PSNR of our method is still larger than that of Ni et al. [5], Zeng et al. [6], and Yang's techniques [4]. Although the PSNR for the proposed method is slightly less than that of Yang and Wang's scheme [7], but the hiding capacity provided by our method is about 2.4 times larger than that provided by their approach [7].

To demonstrate the robustness performance, the marked images generated by the proposed method were tested by various attacks. The extracted watermarks and their bit correct ratio (BCR) are given in Table 3. The size of an input watermark is 254 x 254.The tested marked image is generated by the proposed method using $\tau_r = 8$, $\tau_g = 5$, $\tau_b = 8$ on image Lena. The BCR is defined by

$$BCR = \left(\sum_{i=0}^{ab-1} \overline{w_i \oplus \tilde{w}_i} \Big/ a \times b \right) \times 100\%, \tag{1}$$

where w_i and $\tilde{w}_i$ represent the values of the original watermark and the extracted watermark respectively, as well as the size of a watermark is $a \times b$. The BCR for an extracted watermark is 100% if a marked image is not manipulated. Figure 5 showed the BCR performance of the proposed method under noise-addition, posterized, and color-reduction attacks. Fig. 5(a) indicated that robustness of our method against uniform-noise attacks is better than Gaussian-noise attacks. Fig. 5(b) confirmed the proposed method was capable of resisting from posterized attack with the number of representative colors around 8. Fig. 5(c) demonstrated that even the last 5 bits of the stego-pixels was truncated, the BCR value for our proposed method was about 77%.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3 The marked images generated by the proposed method using various control parameters ($\tau_r/\tau_g/\tau_b$) in host images. (a) Lena (8/5/8), (b) Jet (8/7/7), (c) Peppers (14/0/0), (d) Baboon (4/38/8), (e) House (4/3/7), (f) Sailboat (25/0/0), (g) Splash (16/0/0) and (h) Tiffany (0/0/15)

Table 1 PSNR (dB), payload (bit), control parameters, and number of skipped blocks

Images	PSNR	Payload	Offset ($\tau_r/\tau_g/\tau_b$)	Ratio of skipped and host blocks
Lena	33.88	64,835	8/5/8	1.00%
Jet	33.45	65,146	8/7/7	0.60%
Peppers	33.00	64,545	14/0/0	1.51%
Baboon	24.26	63,612	4/38/8	2.93%
House	36.07	65,292	4/3/7	0.37%
Sailboat	27.89	64,941	25/0/0	0.90%
Splash	31.32	65,484	16/0/0	0.07%
Tiffany	32.37	63,387	0/0/15	3.28%
Goldhill	39.29	65,472	6/0/0	0.10%
Zelda	39.35	65,018	6/0/0	0.79%
Average	33.09	64,773	-	1.16%

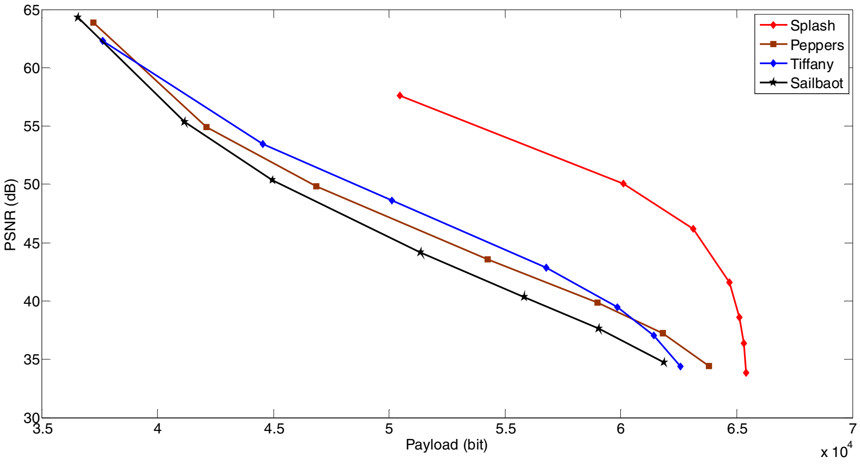
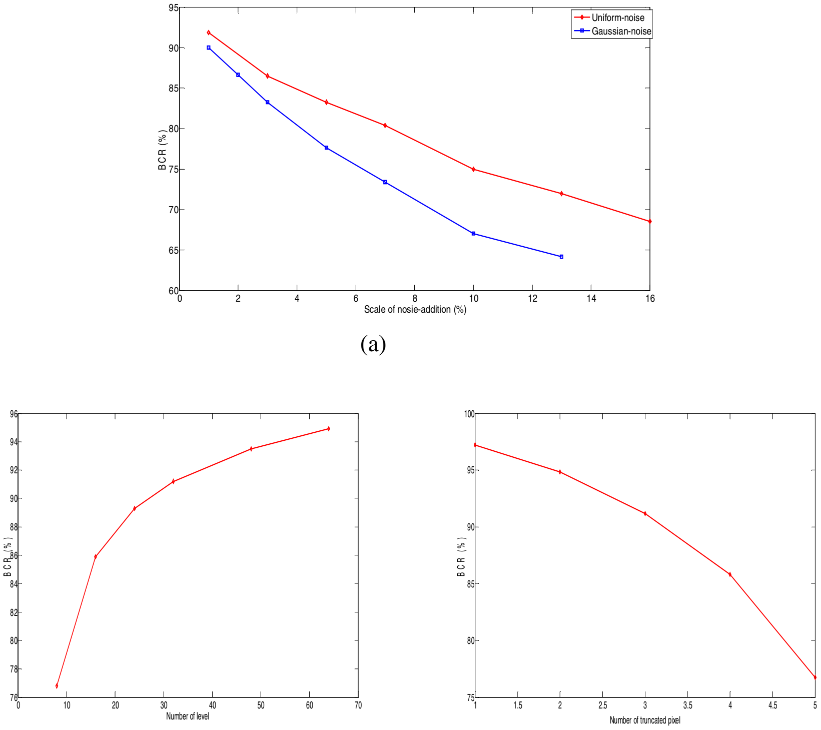


Fig. 4 Trade-off between PSNR and payload for the proposed method in four test images

Table 2 Payload/ PSNR comparison of various methods

Images	Payload/PSNR				
	Ni et al. [5]	Zeng et al. [6]	Yang [4]*	Yang and Wang [7]	Our method
Lena	6,336/ 40.19	16,384/ 38.07	24,576/ 42.18	25,565/ 44.78	62,540/ 42.59
Zelda	4,480/ 40.47	16,384/ 38.09	24,576/ 42.20	27,854/ 43.43	64,268/ 45.49
Goldhill	6,336/ 40.18	16,384/ 38.10	24,576/ 43.50	27,162/ 47.14	64,870/ 45.34
Average	5,717/ 40.28	16,384/ 38.09	24,576/ 42.63	26,861/ 45.12	63,893/ 44.47

*With Yang’s robust watermarking approach.



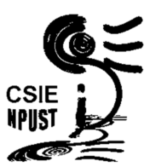
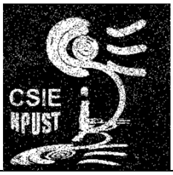
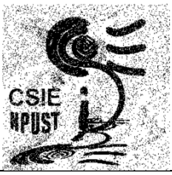
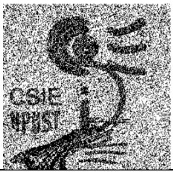
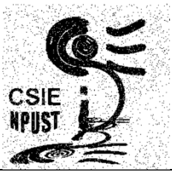
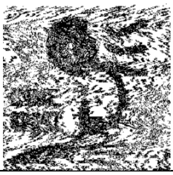
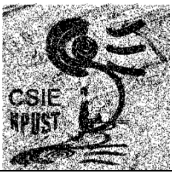
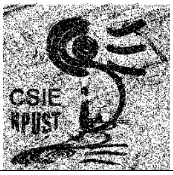
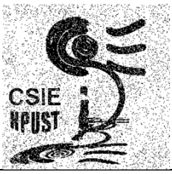
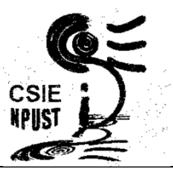
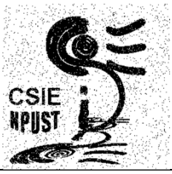
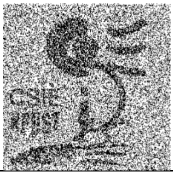
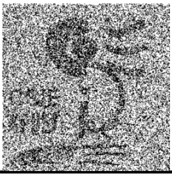
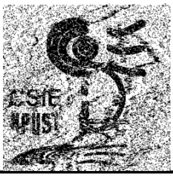
(a)

(b) (c)

Fig. 5 BCR performance of the proposed method under (a) noise-addition, (b) posterized, and (c) color-reduction attacks

Table 3 shows that most extracted watermarks are recognized. Note that BCR for the survived watermark (undergone inversion attack) is only 5.56% and is still recognizable. In addition, the extracted watermarks are recognized when the marked images were compressed by JPEG2000/JPEG with a compression ratio (CR) about 33 and 7, respectively. Similar performance can be found in the marked images which manipulated by brightness, contrast, colour quantization, (edge) sharpening, equalization, noise-addition, pixel-truncation, and winding.

Table 3 The survived watermarks extracted from the marked images which undergone various manipulations.

Attacks	Survived watermarks	Attacks	Survived watermarks
Null Attack BCR = 100%		Inversion BCR=5.56%	
Brightness (150%) BCR=90.19%		JPEG2000 (CR=33) BCR=69.73%	
Contrast (100%) BCR=90.19%		JPEG (QF=70, CR=7) BCR=71.03%	
Color quantization (8-color) BCR=76.79%		Truncation* BCR=76.75%	
Edge Sharpening BCR=91.57%		Sharpening BCR =99.10%	
Equalization BCR=96.22%		Uniform noise (16%) BCR =68.53%	
Gaussian noise (13%) BCR =64.18%		Winding BCR =77.71%	

* The last five-bit of the stego-pixels were purposely truncated.

3 Conclusion

In this paper, a robust reversible watermarking scheme for color images was proposed. By using offset coefficient of a host block, data bits can be effectively embedded in a host image with no noticeable color distortion. Our simulations confirm that the visual quality introduced by the proposed method is good while the payload for the proposed method is better than existing techniques. Moreover, the proposed method shares a certain degree of robustness. The marked images generated by our proposed method are tolerant of manipulations such as brightness, contrast, color quantization, equalization, (edge) sharpening, inversion, JPEG, JPEG2000, noise additions, pixel-truncation, and winding.

References

1. Shih, F.Y. (ed.): Multimedia Security: Watermarking, Steganography, and Forensics. CRC Press, USA (2012)
2. Eielinska, E., Mazurczyk, W., Szczypiorski, K.: Trends in steganography. Communications of the ACM **57**, 86–95 (2014)
3. Noriega, R.M., Nakano, M., Kurkoski, B., Yamaguchi, K.: High payload audio watermarking: toward channel characterization of MP3 compression. Journal of Info. Hiding and Multimedia Signal Processing **2**(2), 91–107 (2011)
4. Yang, C.Y.: Robust watermarking scheme based on radius weight mean and feature-embedding technique. ETRI Journal **35**(3), 512–522 (2013)
5. Ni, Z., Shi, Y.Q., Ansari, N., Su, W., Sun, Q., Lin, X.: Robust lossless image data hiding designed for semi-fragile image authentication. IEEE T. Circuits and Systems for Video Technology **18**(4), 497–509 (2008)
6. Zeng, X.T., Ping, L.D., Pan, X.Z.: A lossless robust data hiding scheme. Pattern Recognition **43**, 1656–1667 (2010)
7. Yang, C.Y., Wang, W.F.: Robust and reversible watermarking scheme for color images based on the RWM norm. In: Annual Conference on Engineering & Information Technology, Tokyo, Japan, March 28-30, pp. 194–203 (2014)

Analysis of Movement Effectiveness in Badminton Strokes with Accelerometers

Wen-Fong Wang, Ching-Yu Yang and Deng-Yi Wang

Abstract The paper focuses on the relevance of three-axis accelerometers and the badminton swing movement coordination and stability. Some of the studies used recording devices or optical instrument to analyze the player's actions, this method is limited by the venue, as well as photography dead, privacy issues, cost, complexity is high. In this paper, the method can effectively solve these problems. The methods uses less computing complexity, achieve effective analysis badminton swing action. After obtaining the three-axis acceleration value, use low pass filtering process for the noise. The characteristic values of different parts of the swing use neural networks to classify the action into error and correct operation provides different feedback for user. We analyzed the data of badminton clear. In the future, we will analysis other strokes in badminton.

1 Introduction

Human motion gesture recognition technology is used in many studies to observe or record people's daily living conditions, especially in some of the need to record the daily activities of patients or people engaged in sports. The technology in addition to doing simple daily activities outside the recording and analysis, has also been

---

W.-F. Wang(✉) · D.-Y. Wang
Department Computer Science and Information Engineering, National Yunlin University of Science and Technology, 123 University Rd., Sec. 3, Douliu, Yunlin 64002, Taiwan
e-mail: wwf@yuntech.edu.tw, m10117019@yuntech.edu.tw

C.-Y. Yang
Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, Makung, Penghu 880, Taiwan
e-mail: chingyu@npu.edu.tw

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_11

used in many different fields, for example: somatosensory operation of gaming, gesture recognition, etc. To study these daily activities, a sensing unit, called inertial measurement unit (IMU), is employed to acquire the movement information. An IMU is an electronic device that measures and reports a craft's velocity, orientation, and gravitational forces, using a combination of accelerometers, gyroscopes, and magnetometers. In addition, as the technology in micro electro-mechanical systems (MEMS) gets a lot of progression, the device such as IMUs can be fabricated in a tiny size to be worn on the human body in an easy way without the limitation of time and locations.

Except the utilization of wearable sensing device to acquire human motions in sports or daily activities, the analysis of pictures from optical cameras or video recorders is also utilized to recognize the human motions. For the video acquisition of motions by cameras and video recorders, a number of reflective sheets must be placed on the relevant parts of the body to collect human body movement signals. Simultaneously, many research works applied the aforementioned method to the motion analysis of professional athletes [2, 3]. In the way of acquiring the body movement signals from cameras and video recorders, the postures of human motions in sports or daily activities can be reconstructed exactly in computers. However, the involved equipment in the aforementioned motion signal acquisition must be established in a fixed location to collect the body movement signals, and it requires a lot of computing resources to complete the motion image analysis work. Therefore, many research works focus on using IMUs for motion signal acquisition in analyzing the body movement of athletes. Especially, the accelerometers appeal to researchers and become the major tool to the motion signal acquisition [4, 5, 6]. In utilizing accelerometers for motion signal acquisition, there are two trends for research. One is to acquire motion signals and then process the signals for later signal classification and the classified result's feedback for training enhancement. The other is to provide beginners a new learning and training method for new sports. In this investigation, we focus on a new way, which provides beginners to practice new sports.

To practice a new sport, it has three primary steps. In the beginning, one can learn the basic movements or sport actions through reading books, watching professional video, or observing the actions of professional athletes. Subsequently, one can adjust his/her movements through the instructions of professionals or coaches or repeatedly viewing other people's action demonstration. Finally, a number of tests such as games, contests, tournaments can be used to examine the effectiveness of learning. In the training process, a beginner needs to be instructed or to view one's video recording in sports in order to get well training effects. Nevertheless, it may happen that the communication between trainers and trainee got problems or one can not be familiar with professional movements; these causes may lead to ineffective practice. In this investigation, we develop a wearable module with an accelerometer inside, which is used to collect the stroke information of badminton players. By the proposed method for motion analysis, the acceleration signals of the badminton strokes can be transformed into readable information for badminton trainee.

The application of accelerometers in the posture recognition of human body is popular recently. A few earlier research works concentrate on recognizing different kinds of human movements such as walking, up/down stairs, siting, standing, falling, etc. The posture recognition of human body can be used to realize the current body movement status [7, 8], but cannot be used to classify those postures with high similarity. However, these research works are still useful for fall detection. For analyzing a long term activity record, a few machine learning algorithms can be applied to classify accurately those highly similar activities [9, 10]. At the same time, some specific activities can be decomposed into small decomposition actions; in this way, other useful information can be obtained [11, 12].

Except basic motion analysis in sports, many sport items related to skilled actions can be analyzed by applying accelerometers to acquire motion signals. For example, in badminton stokes, the body motion's coordination, motion stability, and racquet directionality can also be analyzed by using accelerometers to acquire player's movements. By triaxial accelerometer directionality, one can easily record the movement speed and direction of movement of player's racquet. The action of each movement can be cut into decomposed steps, and then these steps are serialized together through body's coordination and efficiently pass out the power, so that the effectiveness of each action to achieve the best state. Players with good coordination do not need a lot of practice in this regard, they can easily perform each continuous step in perfect way. Another feature is the stability of sport movements. For jogging runners, the more similarity in running speed and stride, the slower the fatigue accumulation will be [6]. In other sports, one can also find a similar situation. The better the ability to players, the more similar in each repetitive action [5, 13, 14].

Acceleration signals generated by triaxial accelerometers are in 3-dimension vector form. A number of studies compute only the signal strength of signal vectors and take that as an important parameter for the identification operation of actions. After determining the target action, one of the axes in the key signal vector is then used to determine the direction of target action carried out [7]. As the amount of signal data obtained in different sampling rates is different. To reduce the amount of signals, the acceleration data are converted into a string, and then use the string to carry out signal analysis [4]. The most important characteristics of the athletic motion stability analysis are to calculate the statistical mean and standard deviation of acceleration signals for the repeatedly actions. If the player is very familiar with the standard operation procedure and urging the difference of swinging a racquet, the related statistics displayed will not be too distinct [13, 14].

In the early research and analysis for badminton, optical cameras were primarily used to capture the player moves, and the purpose of this method is to analyze the swing pictures of good players. In the example of using high-speed cameras to mark human action, the application of this approach requires the establishment of a special site, and players who must be affixed reflective markers. Due to the size of reflective markers are relatively small, and therefore very suitable affixed to any part of the body, so that shows a different racquet swing through the observation and study of reflective markers and obtains the relationship between the related joints and body parts in rotation angle and speed [2]. The common badminton stroke actions such as clear, drop, smash, etc. can be included in the analysis.

Wearable sensing techniques are unrestricted in space, and the incurred cost compared to optical high-speed cameras is much lower. In [15], the authors used some of the MEMS sensor, which is fixed on a badminton racquet for swing movements, to collect information for the analysis of smashing the shuttle. As the racquet swing, they can also analyze which parts of the contact in the racket (racquet line or frame). The results can be the basis of designing and developing new racquets. For the exercise intensity assessment, the general assessment approach is based on the current heart rate to determine the exercise intensity. In [16], the authors find that the acceleration signal may correspond to the heart rate.

There are a lot of motion gesture recognition research in using accelerometers to collect acceleration signals for different parts of the body during exercises. For every skillful action, the acceleration information at each action of outstanding athletes will be very close. In [4], the authors proposed a method in motion transcripts to evaluate the coordination in bat swing actions. They found that the body in the implementation of a bat swing process must be within a certain time and with the correct order in accordance with rotating the relevant parts of the body to perform different decomposition action in order to have a good stroking effect. Through the utilization of motion transcripts, the bat swing process can be analyzed to recognize the body parts in correct rotation order. In [14], the authors used an acceleration sensor to analyze the stability of the tennis player's movements, and their study indicated that in good players' skills, the statistics of acceleration signals in every swing movement are very close, and vice versa for bad skills. In [6], the authors used a number of acceleration sensors to collect jogging acceleration signals in different parts of the human body. From the investigation of these data, the relationship between fatigue and athletic posture changed was analyzed, and their study showed that the running postures of excellent runners can be maintained stable in a normal or fatigue state. On the contrary, the novice runners running posture throughout the jogging process will exhibit an unstable state. Therefore, unstable sport's action is also considered one of the reasons likely to cause fatigue. In [15], a three-axis accelerometer and microphone were mounted in one racquet to record badminton's smash-up action. They calculated the acquired data to get the racquet swing speed and the hitting spot on the racquet, and the results of the experiment showed that the associated swing speed and ball speed in smashing are not highly related.

We learned from the past studies on how to use three-axis accelerometer to collect data during exercises, as well as many different methods and techniques for motion analysis. Until now, it has not yet a motion learning assistant system in badminton for beginner players. In this study, we improved the methods proposed in [4] and [14] and proposed a new method in line with the characteristics of badminton, and through different levels of players to verify the reliability of the system.

In the remaining of this study, Section 2 describes the movements in badminton and their signal processing procedures. It also describes the algorithms of signal acquisition, the features' extraction, and classification. Section 3 shows the experimental results and concludes our work in Section 4.

2 Badminton Actions and Signal Processing

2.1 Descriptions of Badminton Movements

In assessing the effectiveness of badminton swing, we developed a badminton racquet signal acquisition system, which includes an inertial acceleration signal acquisition module equipped with a three-axis accelerometer, as shown in Fig. 1. The racquet is parallel with the Z axis of the triaxial accelerometer, and perpendicular with the Y-axis, when the players perform swing movement. The acceleration signals caused by changes will be recorded in the memory card in the module. There are three common badminton actions, which are clear, drop, and smash. In fact, these three actions have very close decomposed steps. Therefore, the clear action is considered as the most fundamental action in racquet swing.

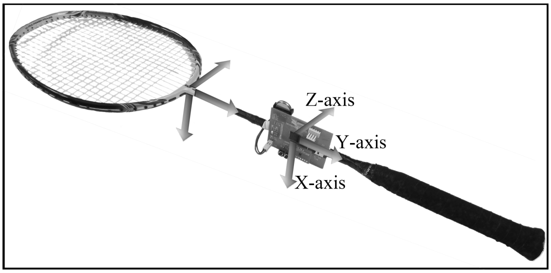


Fig. 1 A badminton racquet integrated with an acceleration signal acquisition module.

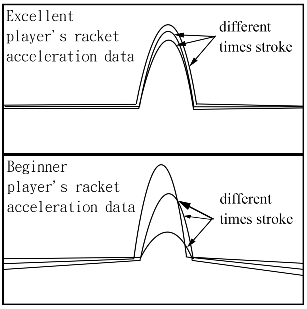


Fig. 2 The acceleration variation for different player in clear actions.

Generally, good players in dealing with the same kind of shuttle pitch by the strokes will be very similar since this is already accustomed to the action for him. So, the acceleration signal of every swing will be very close as shown in Fig. 2. If the players are not familiar with the clear action, his action in every swing will not be consistent. Since the acceleration signals from different players, who are in different

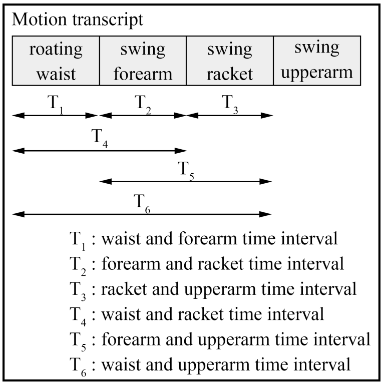

Motion transcript
roating waist
swing forearm
swing racket
swing upperarm
T1
T2
T3
T4
T5
T6
T1 : waist and forearm time interval
T2 : forearm and racket time interval
T3 : racket and upperarm time interval
T4 : waist and racket time interval
T5 : forearm and upperarm time interval
T6 : waist and upperarm time interval

Fig. 3 The timing diagram of the decomposed steps in clear actions from a skillful player.

skill levels, are distinct, we therefore can access the action stability of each player by employing the acceleration statistics from the module on the racquet.

We can understand from the foregoing description. To complete a clear action effectively, we need to swing or rotate sequentially the relevant parts of our body in a timely sequence as shown in Fig. 3. Meanwhile, the difference between the start time of each action can not be too long or short. The aim is to be able to use the entire kinetic chain shot, so that the time difference of each operational sub-steps (e.g. $T_1, \cdots, T_6$) can be considered whether the player action the correct one.

2.2 *Signal Preprocessing*

There are two subsystems developed in our work. The first one is to acquire the acceleration signals in the relevant actions. We called this subsystem the acceleration signal acquisition module, which has the signal storage capability. The acceleration

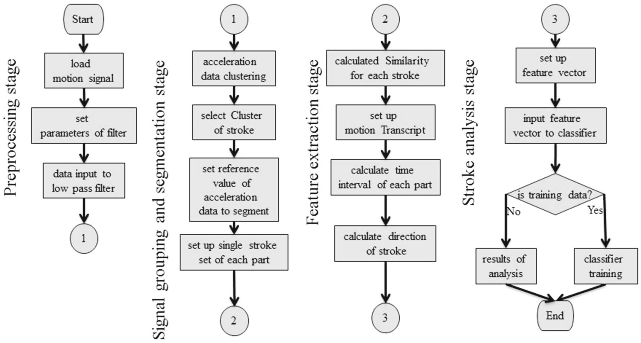

Preprocessing stage
Start
load motion signal
set parameters of filter
data input to low pass filter
1
Signal grouping and segmentation stage
1
acceleration data clustering
select Cluster of stroke
set reference value of acceleration data to segment
set up single stroke set of each part
2
Feature extraction stage
2
calculated Similarity for each stroke
set up motion Transcript
calculate time interval of each part
calculate direction of stroke
3
Stroke analysis stage
3
set up feature vector
input feature vector to classifier
is training data?
No
Yes
results of analysis
classifier training
End

Fig. 4 The workflow of the motion analysis subsystem.

signal data, which are collected from different body parts, can be uploaded to a database in a computer for further processing. The other one is the subsystem for motion analysis. By employing a K-means clustering algorithm, the signals stored in the database are processed for signal segmentation, and then later for signal feature extraction and feature classification. Through the above processing, each action in swing can be evaluated and then feedback to the players.

The main purpose of the motion analysis subsystem is to process the previously established data collection. Through the method proposed in this study, we can quantify the performance of the player's swing and propose amendments in actions. The specific workflow of this motion analysis subsystem is shown in Fig. 4.

2.3 Motion Classification

This study focuses on analyzing three categories badminton characteristics, *i.e.*, the stability, coordination, racket plane direction control of the swing motion, to identify the badminton swing motion effects. After extracting the features in the above badminton swing for clear actions, we need to train three motion classifiers for each of the badminton characteristics as shown in Fig. 5. Since the action of a subject is not necessarily to have the problem in all aspects at the same time, we classify the swing action individually for each of the three categories. As we get the acceleration signals for the stability of $subject_A$, the signals were entered to extract the stability features and then to the $classifier_1$ for the classification result of formal or informal action. In the examples of the characteristics of coordination and racket plane direction control, we employ two separate classifiers, $classifier_2$ and $classifier_3$, for the two different feature's categories. since the subject does not always stroke the shuttle to cause adverse problems. In this way, we can obtain the true shot assessment result. However, the stability is different from the aforementioned two categories since its characteristics is obtained by repeatedly collect the acceleration signals of the subject. The action must be counted more than once at the same time to obtain information on the results. Finally, we use these results to give the assessment of the subject's motion effects

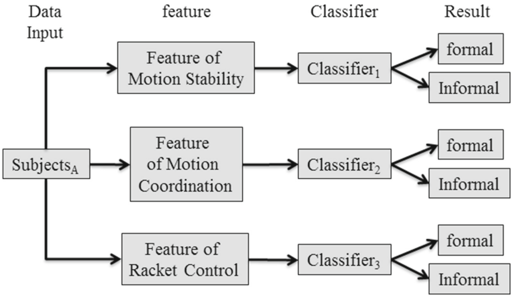


Fig. 5 The workflow of the three motion classifiers.

3 Experimental Results

To conduct the evaluation of our proposed method, fifteen professional players and fifteen common players were recruited from our campus. To identify a player whether his/her performance level is high or not, we collected the acceleration signals for single stroke from the subjects and calculate their signal means. As shown in Figs. 6 and 7, it can be seen for the level difference in professionals and non-professionals.

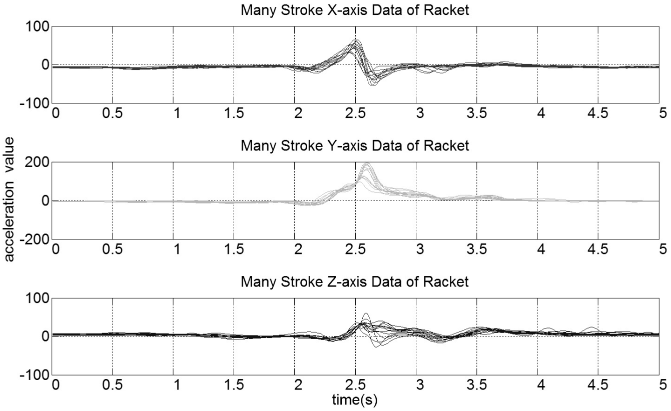


Fig. 6 The signals acquired from good players.

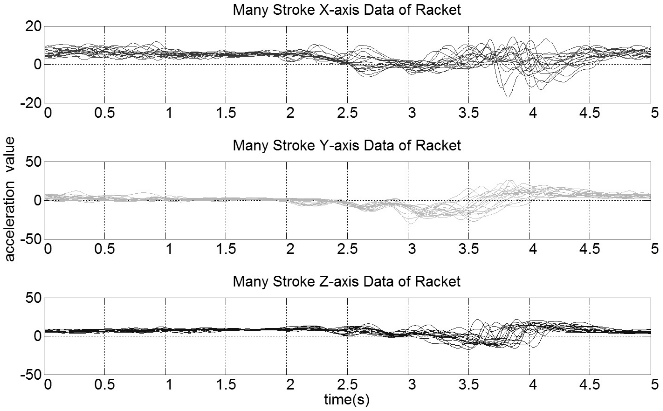


Fig. 7 The signals acquired from common players.

4 Conclusion

We research and development of electronic sports training system can provide a new way to learning movement. Previous studies is limited by special space and special photographic equipment. Our study using wearable sensors to assist players to record stroke motion. We just need a general badminton court and the wearable sensor device is lower cost compared with photographic equipment.

This study proposes a basic motion for badminton clear Analysis System. We use wearable sensors to collect data, and analysis with computer. Finally, give some motion advice for the beginner player. There are 30 players participating in the experiment. We propose three representative assessment indicators, and validate our reliability by F-score.

Badminton motion type in addition to the basic movements, there are many motion type needs to the high level skill. We expect using the acceleration data to build more models badminton motion. In the next study, we will analyze other stroke of badminton motion we expect to develop a complete system capable of training and analyze complete badminton motion.

Acknowledgments This work was supported by the Department of Science and Technology, Taiwan, under contracts MOST 103-2221-E-224-084. At the same time, the authors would like to thank those volunteers in the local campus, who ardently support the conduction of the sport experiments.

References

1. Lee, G.X., Low, K.-S.: A factorized quaternion approach to determine the arm motions using triaxial accelerometers with anatomical and sensor constraints. IEEE Trans. on Instrumentation and Measurement **61**, 1793–1802 (2012)
2. Tsai, C.-L., Hsueh, Y.-C., Pan, K.-M., Chang, S.-S.: Biomechanical analysis of different badminton forehand overhead strokes of Taiwan elite female players. In: ISBS-2008, Seoul, Korea (2008)
3. Salim, M.S., Lim, H.N., Salim, M.S.M., Baharuddin, M.Y.: Motion analysis of arm movement during badminton smash. In: IECBES-2010, Kuala Lumpur, Malaysia (2010)
4. Ghasemzadeh, H., Jafari, R.: Coordination analysis of human movements with body sensor networks: a signal processing model to evaluate baseball swings. Sensors J. **11**, 603–610 (2011)
5. Lai, D.T., Hetchl, H.M., Wei, X., Ball, K., Mclaughlin, P.: On the difference in swing arm kinematics between low handicap golfers and non-golfers using wireless inertial sensors. Procedia Engineering **13**, 219–225 (2011)
6. Strohrmann, C., Harms, H., Kappeler-Setz, C., Trster, G.: Monitoring kinematic changes with fatigue in running using body-worn sensors. IEEE Trans. on IT in Biomed. **16**, 983–990 (2012)
7. Lan, C.C., Hsueh, Y.H., Hu, R.Y.: Real-time fall detecting system using a tri-axial accelerometer for home care. In: Intl. Conf. on Biomed. Eng. and Biotech. (2012)
8. Karantonis, D.M., Narayanan, M.R., Lovell, N., Celler, B.G.: Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Trans. on IT in Biomed. **10**, 156–166 (2006)

9. Wang, Z., Jiang, M., Hu, Y., Li, H.: An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by usingwearable sensors. IEEE Trans. on IT in Biomed. **16**, 691–699 (2012)
10. Gyllensten, I.C., Bonomi, A.G.: Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. IEEE Trans. on Biomed. Eng. **58**, 2656–2663 (2011)
11. Wang, N., Redmond, S.J., Ambikairajah, E., Celler, B.G., Lovell, N.H.: Can triaxial accelerometry accurately recognize inclined walking terrains? IEEE Trans. on Biomed. Eng. **57**, 2506–2516 (2010)
12. Muscillo, R., Schmid, M., Conforto, S., Alessio, T.D.: An adaptive kalman-based bayes estimation technique to classify locomotor activities in young and elderly adults through accelerometers. Medical Engineering & Physics **32**, 849–858 (2010)
13. Guo, Y.W., Liu, G.Z., Huang, B.Y., Zhao, G.R., Mei, Z.Y., Wang, L.: A pilot study on quantitative analysis for table tennis block using a 3D accelerometer. In: Information Technology and Applications in Biomedicine, Corfu (2010)
14. Ahmadi, A., Rowlands, D.D., James, D.A., Ahmadi, A.: Investigating the translational and rotational motion of the swing using accelerometers for athlete skill assessment. In: Sensors, Daegu (2006)
15. Kiang, C.T., Yoong, C.K.: Local sensor system for badminton smash analysis. In: IEEE Intl' Instrumentation and Measurement Tech. Conf., Singapore (2009)
16. Pernek, I., Stiglic, G., Kokol, P.: How hard am i training? Using smart phones to estimate sport activity intensity. In: IEEE Intl' Conf. on Distributed Comput. Syst. Workshops (2012)

A Simple and Non-invasive Measurement Method Important for Conserving Large Animals Underwater

Tadamichi Morisaka, Mai Sakai, Kazunobu Kogi and Hiromitsu Hama

Abstract This paper shows how simple, non-invasive measurement method is important for conserving large animals living underwater. We have studied free-ranging wild Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) around Mikura Island, Japan for about 15 years. We usually observe dolphin behavior and record acoustics underwater including photo-identification of every dolphins. Despite lots of video data, more basic information such as body size have lacked. In order to measure body size of the animals non-invasively, we used commercial measurement software which can measure the distance between two points on 3D photos by commercial 3D camera. Accuracy evaluation tests using a tape measure between 15 and 250 cm were conducted at the field site (Mikura Island) and the errors were within 5 % of the actual sizes. We successfully measured body size of 33 free-ranging wild dolphins. Mean absolute percentage error of the same individuals in different photos was within 4 % for dolphins. The average size of the adult dolphins was 249.6cm, which was the similar to the average actual size of 4 by-caught dolphins of the population (251.7 cm). We also opportunistically measured green turtles (*Chelonia mydas*) during our dolphin survey. Mean absolute percentage error of the same individuals in different photos was within 2 % for turtles. The average straight carapace length (SCL) of the 14 turtles was 48.3 cm

---

T. Morisaka(✉)
Tokai University Institute of Innovative Science and Technology, Shizuoka, Japan
e-mail: chaka@tokai-u.ac.jp

M. Sakai
Faculty of Agriculture, Kinki University, Nara, Japan

K. Kogi
Mikura Island Tourist Information Center, Tokyo, Japan

H. Hama
R&D Center of 3G Search Engine, Incubator, Osaka City University, Osaka, Japan

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing,*
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_12

(37.9 – 67.7 cm), which is much smaller than the adult green turtle length (80 – 111 cm). A simple measurement method is thus important to know the size of the large animals underwater, which indicates animal age and health condition, and are important for conserving these large animals.

1 Introduction

Large animals underwater usually suffers from their extinction. For example, green turtles (*Chelonia mydas*) and loggerhead turtles (*Caretta caretta*) which can be seen around Japan are classified as Endangered in IUCN Red List. The declaration of functional extinction of Yangtze River dolphins (*Lipotes vexillifer*) is still fresh in memory [1]. Such large animals decreased their number by human direct catch, bycatch by fishing gears, and habitat disappearance. Several species, such as Leatherback turtles (*Dermochelys coriaces*) and Vaquita (*Phocoena sinus*) are classified as Critically Endangered in IUCN Red List, and still suffering extinction.

How do we conserve such large animals? It is of course important to decrease direct catch, bycatch, and habitat disturbance of the animal. Also population monitoring is important to find harmful change of the animal population. To monitor population health, basic information should be measured, such as number of the animals in the population, home range or habitat range of the animals, and body size of the animals. If we monitor the trends of these information and detect harmful change, then we are able to move quickly to reduce such harmful effects.

Body size of the animal is the most basic information same as our health examination. Growth rate and/or nutritional status can be measured using accumulation of body size information through several years. In spite of such an important information, it is difficult to measure body size of free-ranging wild large animals, especially underwater. Researchers usually obtain the body sizes of the aquatic large animals from dead animals at beach, stranded animals, by-caught animals, and temporally captured animals. For example, researchers temporally captured bottlenose dolphins in the waters near Sarasota, Florida from 1970 [2] not only to measure body size, but also to take blood sample, tooth (for age estimation), and to acquire an ultrasonic diagnosis for checking pregnancy. Such temporal capturing projects bring basic scientific information in detail, but it needs lots of time, money and hands to capture wild dolphins, and dolphins get stress from capture.

We have studied free-ranging wild Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) around Mikura Island, Japan for about 15 years. At the field site, we basically record dolphin behavior and acoustics underwater including photo-identification of every dolphins using underwater video system [3-9]. In spite of lots of video footage and publications, we do not have basic information about these dolphins, especially body size. How large are they? How large are the neonates at birth? Is there sexual difference in size? What are the growth rate of the neonates, calves and subadults? These parameters are important not only for scientific knowledge, but also to conserve the population as we mentioned earlier.

It is too difficult for us to capture these dolphins temporally because 1) Mikura Island is an oceanic and small island which do not have any bay, beach and shallower water area; 2) commercial dolphin swimming program are conducted on this population during our study period; 3) we do not want to give stress to the dolphins from capture. We thus use commercial measurement software which can measure the distance between two points on 3D photos by commercial 3D camera in order to measure dolphin body size underwater. A simple, handy, light and field-usable system is needed for us biologists rather than accuracy.

We often observe green turtles during our observation of dolphins. No studies except one [10] have conducted on the rare green turtles which are classified as Endangered in IUCN Red List. We opportunistically took 3D photos using same measurement system to know their size, or for age estimation during our survey of dolphins.

2 Materials and Methods

2.1 Subject Animals and Study Area

We collected data from Indo-Pacific bottlenose dolphins residing around the shallower water off Mikura Island, Tokyo (139°36'E, 33°52'N), Japan from 23th, June, 2014 to 2nd, July, 2014. The dolphins have almost all been identified using natural marks on the body by underwater video-identification research since 1994 [11]. They are resident around Mikura Island throughout the year around, and about 120 dolphins were reside in 2014 [11, unpublished data]. We also opportunistically collected data from green turtles when we found them during our observation of dolphins.

2.2 Calibration, Data Acquisition and Analysis

We used simple 3D distance measurement system produced by Applied Vision Systems Corporation, Ibaraki, Japan. This system includes camera calibration program (AVSCalibForFUJI3DCam) and underwater calibration board for camera calibration, and 3D distance measurement software (AVS-Measure3D-F). We used FinePix Real 3D W3 (Fujifilm Corporation, Tokyo, Japan) with underwater housing (WHF-3DW3; NTF Corporation, Kanagawa, Japan) for taking 3D photos of the animals. The principle of the distance measurement, and calibration is shown in detail in the webpage of Applied Vision Systems Corporation (http://avsc.jp/technology/technology_principle.html; in Japanese).

We calibrated camera system (camera with underwater housing) at the Port of Mikura Island using the software and the calibration board. We took pictures from various angles and the software provided the calibration data. This calibration data was embedded in 3D distance measurement software. We then conducted accuracy evaluation tests using tape measure from 15 to 200 cm, and using a 250 cm rope at the Port of Mikura Island from various angles and distances.

We took 3D photos using the system which attached to our underwater video recording system (Fig. 1). When we spotted dolphins from the dolphin-swimming boat, we slowly approached the group and enter the water with the system. We used only fins, snorkels and masks without scuba diving equipment. If the dolphins came appropriate distance where whole dolphin body can be record within the camera flame, we took the photos. If the turtles were found at that time, we took the photos.

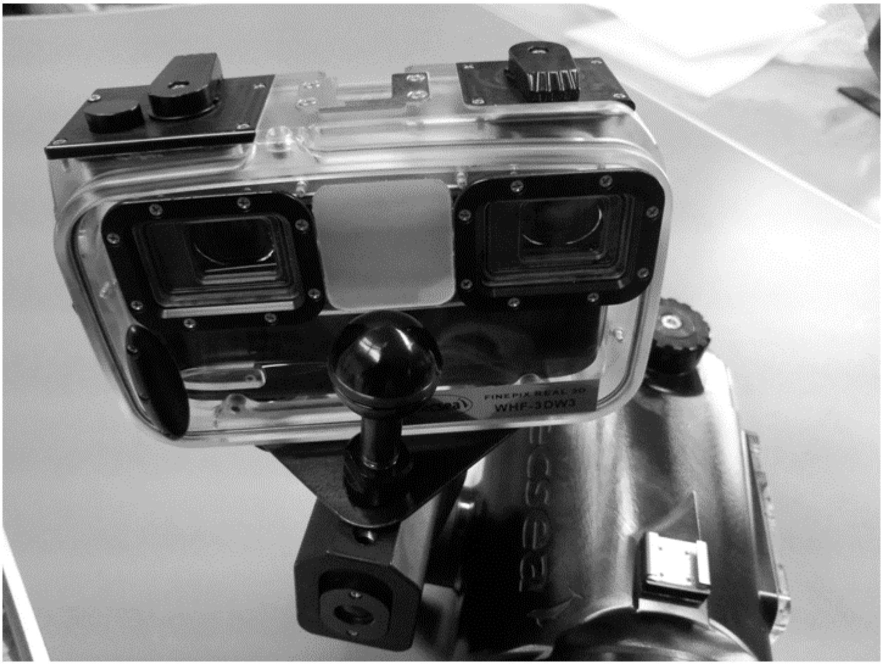


Fig. 1 The camera system attached to the underwater video recording system

We used 3D distance measurement software to measure body size of the animals. We first identified dolphins using ID catalogue which was provided by Mikura Island Tourist Information Center. Then we measured standard length (straight line from the tip of the upper jaw to the fluke notch) of the dolphin body. We repeatedly measured (more than three times) dolphin body size of each pictures. After discarding obvious abnormal value, we averaged measurement values. Mean absolute percentage error of the same individuals in different photos was calculated to know the accuracy of the system using following equation;

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{At-Ft}{At}\right| \times 100 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where M is the mean absolute percentage error, *At* is the actual value of the body size, and *Ft* is the average body size of each individuals. If the body seemed to be bend, we did not use that picture. For the body size of the green turtles, we measured straight carapace length (SCL) which was measured from the anterior point at midline to the posterior notch at midline between the supracaudals [12].

3 Results

3.1 Accuracy Evaluation Test

Accuracy evaluation tests using tape measure between 15 and 250 cm at the Port of Mikura Island revealed that the measuring errors were within 5 % of the actual sizes (Fig. 2). Error % tended to become larger when the actual size became larger, but did not change along with the angle of the tape measure relative to the camera.

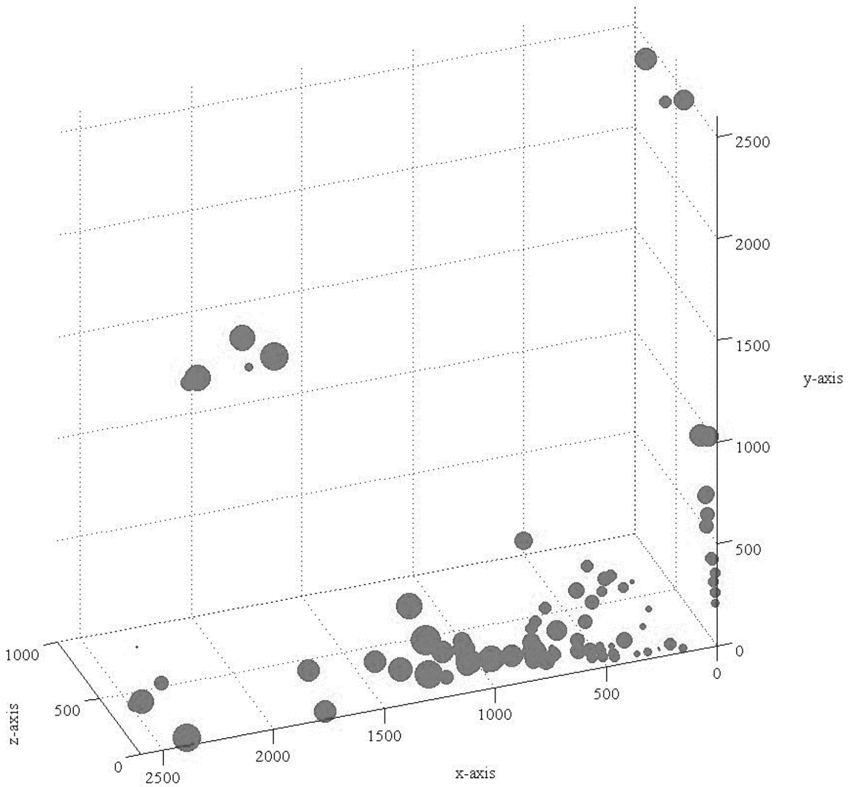


Fig. 2 Three dimensional error distribution of accuracy evaluation test. Circles indicate error % (relative value) to the actual size, and circle locations indicate the end of the tape measure relative to the camera (another end is always the origin).

3.2 Dolphin Body Size

We successfully measured standard length of the dolphin body from 33 individuals. Seventeen dolphins were adult dolphins over 13 years. There was no significant difference between the estimated body size of the adult dolphins over 13 years old (249.6 ± 10.7 cm) and 4 by-caught adult dolphins over 13 years old from same population (251.7 ± 5.5 cm) (Wilcoxon/Kruskal-Wallis test; λ^2=0.07, p=0.79). Mean absolute percentage error of the 6 same individuals in different photos was within 4 % for dolphins.

3.3 Turtle Size

We successfully measured SCL from 14 individuals with 48.3 ± 10.2 cm (37.9 – 67.7 cm). Mean absolute percentage error of the 3 same individuals in different photos was within 2% for turtles.

4 Discussion

We successfully measured body sizes of free-ranging wild Indo-Pacific bottlenose dolphins and green turtles using simple and non-invasive underwater system within 5 % error to the body size. Although the error rate should be smaller, it still worth measuring body size of the free-ranging wild large underwater animals. The system was small, light and easy to use for the underwater researchers, and can use less than 40 m depth. In future, the accuracy of this system should be increased after finding correction equation per distance and actual size.

No significant difference was found between the measured adult size and the actual size of by-caught adults in dolphins. It suggested that the estimation was not abnormal but good. Wang et al. (2000) showed that Indo-Pacific bottlenose dolphin around Taiwan and other place were between 140 and 268 cm [13]. Our data were almost within this range, suggesting again that our estimation was not abnormal. We have no ecological information about the green turtles around Mikura Island so far. Minimum SCL size for the adult green turtle was reported 81 cm [14], so the turtles around Mikura Island seems immature individuals.

A simple measurement method is thus important to know the size of the large animals underwater, which indicates animal age and health condition, and are important for conserving these large animals. It is important to monitor such body sizes of the animals in order to detect harmful change, and will result quick move to reduce such harmful effects on the endangered animals.

Acknowledgment This research was supported by JSPS KAKENHI Grant Number 23310166 and partially 23220006. We are grateful to Yoshinori Hirose, Shingo Hirose, Michio Kurimoto, Nobuo Hirose, Masakazu Hirose, Soji Hirose, Yoshihiko Hirose, Ichiro Kurimoto, Kohei Omori, Naoki Hirose, Mitsuki Inoue, Shigemasa Hirose and Keiji Kato for permission to board their boats without a fee. We also thank Toru Hishii, Mitsuko Hirose, Akihiko Hirose, Masamichi Kuroda, Isoko Kokatsu, Keiko Hirose, members of the Mikura Island Tourist Information Center, the MIDO team, and people living in Mikura Island for their support for conducting this study.

References

1. Turvey, S.T., Pitman, R.L., Taylor, B.L., Barlow, J., Akamatsu, T., Barrett, L.A., Zhao, X., Reeves, R.R., Stewart, B.S., Wang, K., Wei, Z., Zhang, X., Pusser, L.T., Richlen, M., Brandon, J.R., Wang, D.: First human-caused extinction of a cetacean species? Biology Letters **3**, 537–540 (2007)
2. Read, A.J., Wells, R.S., Hohn, A.A., Scott, M.D.: Patterns of growth in wild bottlenose dolphins, *Tursiops truncatus*. Journal of Zoology **231**, 107–123 (1993)
3. Morisaka, T., Shinohara, M., Nakahara, F., Akamatsu, T.: Effects of ambient noise on the whistles of Indo-Pacific bottlenose dolphin population. Journal of Mammalogy **86**, 541–546 (2005)
4. Morisaka, T., Shinohara, M., Nakahara, F., Akamatsu, T.: Geographic variations in the whistles among three Indo-Pacific bottlenose dolphin *Tursiops aduncus* populations in Japan. Fisheries Science **71**, 568–576 (2005)
5. Sakai, M., Hishii, T., Takeda, S., Kohshima, S.: Laterality of flipper rubbing behaviour in wild bottlenose dolphins (*Tursiops aduncus*): caused by asymmetry of eye use? Behavioural Brain Research **170**, 104–210 (2006)
6. Sakai, M., Hishii, T., Takeda, S., Kohshima, S.: Flipper rubbing behaviors in wild bottlenose dolphins (*Tursiops aduncus*). Marine Mammal Science **22**, 966–978 (2006)
7. Sakai, M., Morisaka, T., Kogi, K., Hishii, T., Kohshima, S.: Fine-scale analysis of synchronous breathing in wild Indo-Pacific bottlenose dolphins (*Tursiops aduncus*). Behavioural Processes **83**, 48–53 (2010)
8. Morisaka, T., Sakai, M., Kogi, K., Nakasuji, A., Sakakibara, K., Kasanuki, Y., Yoshioka, M.: Spontaneous ejaculation in a wild Indo-Pacific bottlenose dolphin (*Tursiops aduncus*). PLoS ONE **8**, e72879 (2013)
9. Morisaka, T., Sakai, M., Kogi, K., Shinke, T., Zin, T.T., Hama, H.: A mobile measurement system for sound source level of wild dolphin whistles. Memoirs of the Faculty of Engineering, Osaka City University **54**, 25–29 (2013)
10. Morisaka, T., Sakai, M., Kogi, K., Kubota, S., Kamezaki, N.: A green turtle (*Chelonia mydas*) feeding on a jellyfish around Mikura Island, Tokyo, Japan. Umigame Newsletter Japan **89**, 15–16
11. Kogi, K., Hishii, T., Imamura, A., Iwatani, T., Dudzinski, K.M.: Preliminary demographic parameters and reproductive rates of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) around Mikura Island, Japan. Marine Mammal Science **20**, 510–526 (2004)
12. Bolten, A.: Techniques for measuring sea turtles. In: Research and Management Techniques for the Conservation of Sea Turtles, pp. 1–5. IUCN/SSC Marine Turtle Specialist Group Publication, No. 4, Wachington, DC, USA (1999)
13. Wang, J.Y., Chou, L.-S., White, B.N.: Differences in the external morphology of two sympatric species of bottlenose dolphins (genus *Tursiops*) in the waters of China. Journal of Mammalogy **81**, 1157–1165 (2000)
14. Ishihara, T.: Life cycle. In: Natural History of Sea Turtles in Japan, pp. 57–83. University of Tokyo Press, Tokyo (2012)

Characterization of Malware Detection on Android Application

Chit La Pyae Myo Hein and Khin Mar Myo

Abstract Mobile malware performs malicious activities like stealing private information, sending message SMS, reading contacts and can even harm by exploiting data. Malwares are spreading around the world and infecting not only for end users but also for large organizations and service providers. Malware classification is a vital component and works together with malware identification to prepare the right and effective malware antidote. Not only malware classification but also malware characterization is also important to inform the user and install the malware application, because user are not aware to install the several application of their device. Android application requires several permissions to work. An essential step to install an Android application into a mobile device is to allow all permissions requested by the application. An application is being installed; the system prompts a list of permissions requested by the application and asks the user to confirm the installation. Although Google announced that a security check mechanism is applied to each application uploaded to their market, the open design of the Android operating system still allows a user to install any application downloaded from an unsafe source. Nevertheless, the permission list is still the minimal defense for a user to detect whether an application could be harmful. Therefore in my research work, characterization is still needed to be completed. Malware characterization will also be performed to identify different features from the manifest file. The proposed system can be used to improve the efficiency of Android permission system for informing the user about the risks of Android permissions and applications.

Keywords Android security · Malware · Smartphone

C.L.P.M. Hein(✉) · K.M. Myo
University of Computer Studies, Mandalay, Mandalay, Myanmar
e-mail: {chitlapyae86,kmmyo.ag}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_13

1 Introduction

Firstly we made a lot of reviews on mobile malware analysis, classification and detection areas. After that, we found that there are three main approach to classify and detect malwares: (i) Static Approach (Code based), (ii) Dynamic Approach (Run time Based) and (iii) Hybrid (Static and Dynamic) Approach. More efficient and effective method for Static-Based malware detection is still needed to find out, because static approach has the advantages of less cost rather than dynamic approach. Static-Based malware detection of the risk assessment method proposed and studies permission combinations as risk signals. The analysis of risk signals, however, is based on a malware application sample from the Cantagio-Dump. According to the experiment on different categories, the results show that our proposed permission request characterization is applicable as a light weight approach.

The main goal of this system is to study tools and techniques to assist security analysts and final users in the analysis of entrusted applications for smart devices and automate the identification of smart malware.

To achieve this goal, we will focus in the following three general objectives:

- Study the evolution and current state of malware for smart devices, as well as recent progresses made to categorize and detect it.
- Develop techniques aiming at better categorize malware in large scale software markets, with particular emphasis on intelligent instruments to automate parts of the analysis process.
- Facilitate the analysis of complex smart malware in scenarios with a constant and large stream of applications on target. Examples of such sophistication include malware targeting user-specific actions, malware hindering detection with advance obfuscation techniques, or malware exploiting the battery limitations of current devices, to name a few.
- Demonstrate that exploiting static features for characterization of malware in terms of three key factors: pursued goals and associated behaviors; distribution and infection channels; and privilege acquisition strategies.

Therefore, in this work is to explain the criteria for characterization, we need to process characterize different features from the manifest file. The behavior of malware features is using patterns and tests will do by applying the defined patterns and review the results of the malware detection. Android security model is a permission-based system that provides controlled access to various system resources. Android security plays a vital role in today's mobile world. Android has a large number of permissions restricting access to advanced functionality on devices, only a small number of these permissions are actively used by developers. Android permission system denies access to functionality or other applications

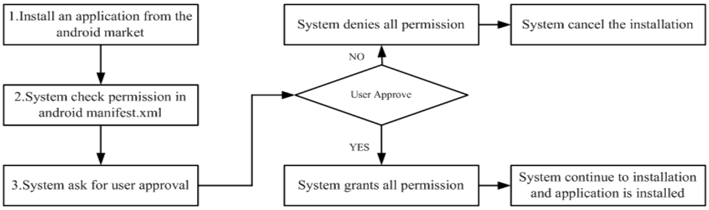

1.Install an application from the android market
2.System check permission in android manifest.xml
3.System ask for user approval
User Approve
NO
YES
System denies all permission
System cancel the installation
System grants all permission
System continue to installation and application is installed

Fig. 1 Android security model of permission request to user

that could negatively impact that is installed on the device as shown in Figure [1]. Examples:

- Sending message or making phone calls (which may cost to the user);
- Keeping the device screen on or accessing the vibrator (which may drain the battery);
- Reading the user's address book (which may violate the user's privacy).

The number of version defined in Android platform tends to increase, and the API- level set of permissions is the most frequent and continues to grow. There were 112 Android permissions in the first widely-used release (API level 8); the number of permissions has grown to 145 in the most current release (API level 19) as shown in Table [1].

Table 1 Version of the Android Platform

Version name	Version number	API Level	Total Permissions
KitKat	4.4	19	145
Jelly Bean	4.3	18	134
	4.2	17	130
	4.1	16	130
Ice Cream Sandwich	4.0.3	15	124
	4.0	14	122
Honeycomb	3.2	13	117
	3.1	12	116
	3.0	11	116
Gingerbread	2.3.4	10	115
	2.3.3	9	115
Froyo	2.2	8	112

Furthermore, the total permissions is always the largest group across all API levels, e.g., Jelly Bean and its' version number 4.2 out of 130 permissions in API level 17, and is still growing. To facilitate application construction, the Android platform provides a structure of application developers. The classify consists of Android packages and classes, attributes for declaring and accessing resources, a set of Intents, and a set of permissions that applications can request. This classify is accessible to apps via the Android application programming interface (API). The Android platform has under gone many changes since its inception in 2008, and each major release forms a new API level. In this paper we studied all major API levels, from level 8 to level 19. With each API upgrade, the older version change the new version and permission are added the developer to inject the malware permission, so that existing applications are limit the total number of permission.

There are many research concerning with Malware Detection, it is still needed to perform. New malicious malware patterns are made by developer. New permission request types are allowed in every new Android Version. It is more difficult to determine whether the application is genuine or not. Static approach detection is still being worked among researchers because the low cost manner of static approach.

The rest of the paper is organized as follows. Section 2 presents the literature review .Section 3 explain the permission request on android application, detailing the process of building the application to collect and give information about malware detection system. Section 4 presents the experimental results of the malware detection testing and evaluation methods. Section 5 experimental validation and conclusion gives possible future work to express limitations of the system proposed.

2 Literature Review

Many researchers propose complicated extensions to fortify the Android's security framework. They mainly focus on protecting the user data and mitigating some types of privilege escalation attacks. This section recent some of the most well-known approaches to extract the malware list in android technologies.

To avoid the malware attack Google itself provide cloud services such as Google Play which is collection of services that allow user to discover, install or buy application on their android device in addition it provide community review application license verification, application security scanning and other security services [6], Security Mechanisms need to be made concrete as per the new trends. It is seen the refinements that need to be done in current security models. Hence next work could be to build a security mechanism that could take care of issues that we've discussed that would refine the permission mechanism to avoid monitor unnecessary access to the resources. Many new frameworks like MADAM [14, 15] a multi-level anomaly detector for android malware which concurrently monitors Android at the kernel level and user-level to detect real malware infections help in making android secure .It is clear that user involvement is greater & important in android security than other mobile phone platform hence user should be made aware of the basic possible threats at the beginning of his android experience further various frameworks like KIRIN [16], SCanDroid [17] , Saint[18] , Apex[19] also purposed.[14]

Alternative research has focused on using machine learning techniques to identify malware. [4]Sanz et al. (2012) applied several types of classifiers to the permissions, ratings, and static strings of 820 applications to see if they could predict application categories. They applied this by using the category scenario as a stand-in for malware detection. [1] Shabtai et al.(2010) similarly built a classifier for Android games and tools, as a proxy for malware detection.[12] Ryo Sato, Daiki Chiba and Shigeki go to proposes a new method for detecting Android malware by analyzing only manifest files based on malware score. [15] Zhou et al. (2012) found real malware in the wild with DroidRanger, a malware detection system that uses permissions as one input. DroidMat (2013) focuses on using attributes of the manifest to trace API calls requiring permissions. [11] N. Peiravian and X. Zhu (2013) propose a rule-based security mechanism designed to prevent malware at install-time. [14] Explores the use of machine learning algorithms for malware detection using permissions and API calls. The studies in Mila (2009) and android (2015) Android Content License "URL" www.source.android.com /license.html retrieved focus on efficient, scalable, and accurate malware detection in large Android markets.

According to Wu, Mao, Wei, Lee and Wu. (2012) there have been a lot of methods and techniques for feature selection. Most of the techniques are based on machine learning technique. [11] Meanwhile other papers Peiravian and Zhu (2013) Shabtai, Fedeland Elovicitried (2010) out another light weight approaches to classify mobile malware. Their approach [11] [3] is only based on manifest file analysis rather than applying machine learning algorithms.

3 Android Permission System

The Android system uses several methods to secure the devices of the users. We will describe the security features that affect applications directly, which are the features that are relevant for malicious applications to attempt to defeat or circumvent. Malware is one of the most important issues in Android. Mobile malware performs malicious activities like stealing private information, sending sms, reading contacts and can even harm by exploiting data.

The permission system [36] is comprehensive and provides a good framework for determining what resources an application will have access to once it has been installed on a device. All the applications for Android devices, including malware, are encapsulated in a specific way known as APK (Application Package File). This format is used for the installation and distribution of applications for this mobile platform. These permissions are stored in a file called manifest.xml found inside the APK file of the application, and cannot be changed after the application is installed. The structure of APK files is shown in Figure [2].

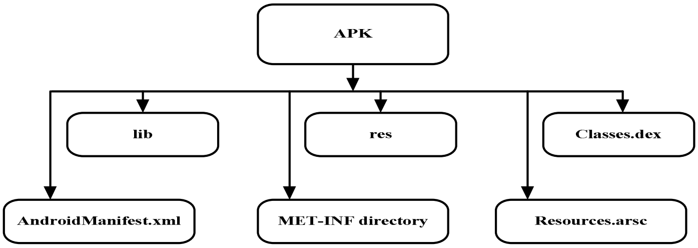


Fig. 2 Directory Structure of an unzipped APK files

Permissions have different danger levels depending on the functions they allow the application to perform and are consequently classified in protection level groups. Likewise, through this attribute, it is possible to determine which applications have access to the permission:

Risk1: They pose a low-risk factor and typically only affect the application's scope. Risk1 permissions are granted by the system automatically without explicit approval of the user.

Risk2: They are medium-risk permissions that allow costly access to services. The permissions can be granted by the user during installation. If the permission request is denied, then the application is not installed.

Risk3: They are high-risk permissions are only granted if the requesting application is signed by the same developer that defined the permission. Risk3 permissions are useful for restricting component access to a small set of applications trusted and controlled by the developer.

The big problem is that groups can contain both normal, basic permissions as well as more dangerous permissions. For example:

- Location: An app that asks for your approximate, network-based location can now gain permission to track your exact location with your device's GPS.
- Photos/Media/Files: An app that needs to read the contents of your USB storage or SD card can now format your entire external storage device.
- Camera/Microphone: An app that has permission to take pictures and videos (for example, a camera app) can now gain the permission to record audio. The app could listen to you when you use other apps or when your device's screen is off.

Table [5] gives some examples of permissions; Android OS defines about several permissions which are specified as static analysis members in the android. Manifest. permission class.

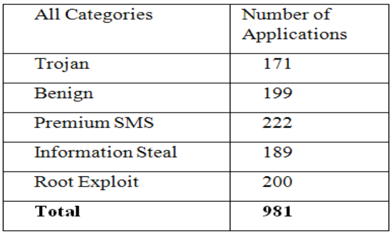
4 Experimental Results

Malwares are spreading around the world and infecting not only for end users but also for large organizations and service providers. In response to the rapid distribution of Android malware, there is a real need for tools that can automatically detect malicious applications. This system describes detecting malware in the Android platform by performing static analysis on the permission based framework in Android platform.

A. Experimental Set Up

We determined the number of samples application for each category in ContagioDump and available at http://cgi.cs.indiana.edu/~nhusted/dokuwiki/doku.php?id=datasets, we randomly selected the applications. The number of samples obtained for each category is shown in Table [2].

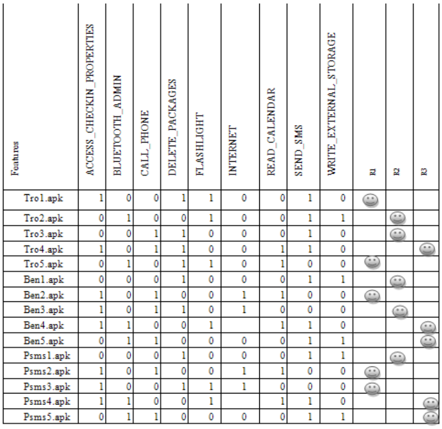
Table 2 Experimental Dataset of Cantagiodump



All Categories	Number of Applications
Trojan	171
Benign	199
Premium SMS	222
Information Steal	189
Root Exploit	200
Total	**981**

B. Feature Extraction

The most malware applications gain access over the mobile phone system through the vulnerable "permissions" feature and hence, we use this particular feature as an indication of a specific malicious app. Besides, we can also extract the API calls from the source code of the app, found in "classes.dex" file of the APK as it carries substantial semantic level information about the app behavior [24]. We can also extract the strings from the app because it has been successfully used in identifying PC malware [25]. The experiment is conducted in a constrained virtualized environment where the benign and malicious application files are first downloaded. These files are available for download in .apk format from various sources. Although these files have the .apk extension, they are simply ZIP files. They are decompressed to extract the content inside the apk package. We use an open source apktool in order to extract features and conduct our analysis. Consequently, we extract the permissions required by the application in order to run it, using the aapt dump command. The android platform facilitates to recognize these permissions as they are stored in a configuration file called the “AndroidManifest.XML”. This file contains all the requirements including permissions, pertaining to the execution of the application. Other requirements may include the minimum OS version required for the application to run, etc. Figure [3] illustrates the process of feature extraction carried out in order to conduct our initial analysis of the idea of feature’s character.



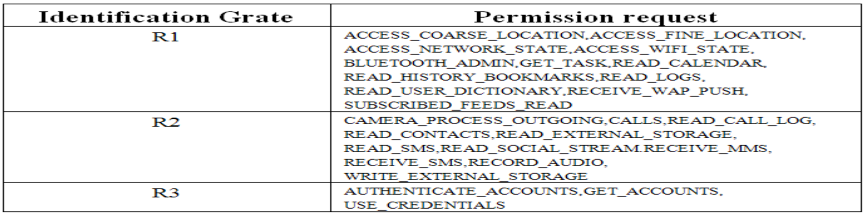
Features	ACCESS_CHECKIN_PROPERTIES	BLUETOOTH_ADMIN	CALL_PHONE	DELETE_PACKAGES	FLASHLIGHT	INTERNET	READ_CALENDAR	SEND_SMS	WRITE_EXTERNAL_STORAGE	R1	R2	R3
Tro1.apk	1	0	0	1	1	0	0	1	0	☺		
Tro2.apk	0	1	0	0	1	0	0	1	1		☺	
Tro3.apk	0	0	1	1	0	0	0	1	0		☺	
Tro4.apk	1	0	1	1	0	0	1	1	0			☺
Tro5.apk	0	1	0	1	1	0	1	0	0	☺		
Ben1.apk	0	0	0	1	0	0	0	1	1		☺	
Ben2.apk	1	0	1	0	0	1	1	0	0	☺		
Ben3.apk	1	0	1	1	0	1	0	0	0		☺	
Ben4.apk	1	1	0	0	1		1	1	0			☺
Ben5.apk	0	1	1	0	0	0	0	1	1			☺
Psms1.apk	0	0	0	1	0	0	0	1	1		☺	
Psms2.apk	1	0	1	0	0	1	1	0	0	☺		
Psms3.apk	1	0	0	1	1	1	0	0	0	☺		
Psms4.apk	1	1	0	0	1		1	1	0			☺
Psms5.apk	0	1	1	0	0	0	0	1	1			☺

Fig. 3 The process of feature extraction

C. Feature of Permission

The static analysis of applications is often to classify an application as malicious or benign. In classification features are used to make informed decisions. Permissions allow an application to access potentially dangerous API functionality. Many applications require several permissions to function properly. These permissions must be listed explicitly in the application’s AndroidManifest.xml file and accepted by the user during installation. Permissions can be considered as features that describe the functionalities of applications, or the application’s behavior indicating its attempt to interact with the system, the data or other applications. The malware applications are basically different from Cantagio on requesting different permissions, i.e., having different values of permission/feature variables. Defining a class variable that indicates the label of an application, benign or malicious, the risk of granting permission can be evaluated by measuring the relevance between this permission/feature variable and the class variable. A strong correlation of them indicates the high risk of granting this permission. Measuring the relevance of a feature and class variables is known as feature ranking in machine learning, which has a goal of selecting the most informative features and improving the performance of learned models.

Table 3 Identification Grate of Risk Level

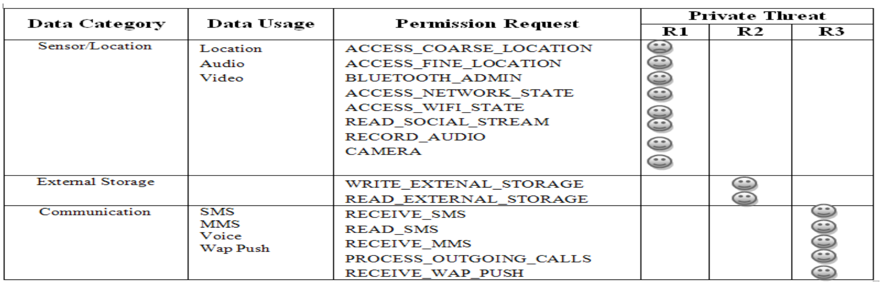


Identification Grate	Permission request
R1	ACCESS_COARSE_LOCATION,ACCESS_FINE_LOCATION, ACCESS_NETWORK_STATE,ACCESS_WIFI_STATE, BLUETOOTH_ADMIN,GET_TASK,READ_CALENDAR, READ_HISTORY_BOOKMARKS,READ_LOGS, READ_USER_DICTIONARY,RECEIVE_WAP_PUSH, SUBSCRIBED_FEEDS_READ
R2	CAMERA_PROCESS_OUTGOING,CALLS,READ_CALL_LOG, READ_CONTACTS,READ_EXTERNAL_STORAGE, READ_SMS,READ_SOCIAL_STREAM.RECEIVE_MMS, RECEIVE_SMS,RECORD_AUDIO, WRITE_EXTERNAL_STORAGE
R3	AUTHENTICATE_ACCOUNTS,GET_ACCOUNTS, USE_CREDENTIALS

Table [3] show the requested permissions in the benign application and malware datasets. Risk level of identification grate R2: CAMERA is the most frequently used permission by both the benign applications and malware. There are many reasons to request permission for picture access: some of the applications need to log in; some are designed to use internet like browsers and email clients; some need to load advertisement etc. As a result, Camera-related permissions, such as ACCESS NETWORK STATE and ACCESS WIFI STATE, become very popular. Another set of widely used permissions are location related ones such as ACCESS FINE LOCATION and ACCESS COARSE LOCATION for location based services as defined on grate level of R1,etc.

Another observation is that some applications of permissions are requested by such malware applications in the Cantagio dataset. Malware are more favor of changing the settings and use money-related services such as short message service (SMS). Changing settings, especially changing the network settings, generally is the first step before a malware performs any malicious activity. Sometimes malware even try to kill background processes, which could help them avoid being detected by anti-virus applications. We can see that the usage pattern of SMS related permissions is quite different between the benign applications and the malware applications and many malware applications attempt to request SMS related permissions. SMS is also a risky permission of private threat (Risk3) that is more likely requested by malware applications. Describe the data usage of SMS include the data category of communication as shown in Table [4].

Table 4 Categorization of Risky Permission



Data Category	Data Usage	Permission Request	Private Threat		
			R1	R2	R3
Sensor/Location	Location Audio Video	ACCESS_COARSE_LOCATION ACCESS_FINE_LOCATION BLUETOOTH_ADMIN ACCESS_NETWORK_STATE ACCESS_WIFI_STATE READ_SOCIAL_STREAM RECORD_AUDIO CAMERA	☹ ☺ ☺ ☺ ☺ ☺ ☺ ☺		
External Storage		WRITE_EXTENAL_STORAGE READ_EXTERNAL_STORAGE		☺ ☺	
Communication	SMS MMS Voice Wap Push	RECEIVE_SMS READ_SMS RECEIVE_MMS PROCESS_OUTGOING_CALLS RECEIVE_WAP_PUSH			☺ ☺ ☺ ☺ ☺

Table 5 Detail Description of Application Permission Request

Request Application Permission	Official Description	Details
Send SMS or MMS *Services that cost you money* ***android.permission.SEND_SMS*** ***HIGH*** *Risk*	Allows an application to send SMS messages.	This permission is of high importance. This could let an application send an SMS on your behalf, and much like the phone call permission, it could cost you money by sending SMS to for-pay numbers. Certain SMS numbers work much like 1-900 numbers and automatically charge your phone company money when you send them an SMS.
Modify/delete SD card contents *Your personal information* ***android.permission.WRITE_EXTERNAL_STORAGE*** ***MEDIUM*** *Risk*	Allows an application to write to external storage	This permission is of high importance. This will allow applications to read, write, and delete anything stored on your phone's SD card. This includes pictures, videos, mp3s, documents and even data written to your SD card by other applications. However, there are many legitimate uses for this permission. Many people want their applications to store data on the SD card, and any application that stores information on the SD card will need this permission. You will have to use your own judgment and be cautious with this permission knowing it is very powerful but very, very commonly used by legitimate applications. Applications that typically need this permission include (but are not limited to) camera applications, audio/video applications, document applications.
Read Contacts *Development tools / Your personal info* ***android.permission.READ_CONTACTS*** *LOW Risk*	Allows an application to read the user's contacts data.	This permission is of high importance. Unless an app explicitly states a specific feature that it would use your contact list for, there isn't much of a reason to give an application this permission. Legitimate exceptions include typing or note taking applications, quick-dial type applications and possibly social networking apps. Some might require your contact information to help make suggestions to you as you type. Typical applications that require this permission include: social networking apps, typing/note taking apps, SMS replacement apps, contact management apps.

5 Conclusion

As mobile devices begin to see an increasing volume of malicious applications, the ability of application markets to identify and remove such applications in a favorable manner will be lost. We respond to this problem by developing detection of Android malware characterization. The advantage of this category is that it uses only manifest files to detect malware. Manifest files are required in all Android applications, and thus, the proposed system is applicable to all Android applications. Future work will emphasize testing of already tested malware applications to deduce the characters of their performance. Testing of other major applications has been done in the Android Market and discovers additional mobile device vulnerabilities.

References

1. Fuchs, A., Chaudhuri, A., Foster, J.: Scandroid: Automated security certification of android applications. Manuscript, Univ. of Maryland, May 7, 2010
2. Shabtai, A., Fledel, Y., Elovici, Y.: Automated static code analysis for classifying android applications using machine learning. In: Proceedings of the 2010 International Conference on Computational Intelligence and Security (CIS) (2010)
3. Felt, A.P., Greenwood, K., Wagner, D.: The effectiveness of application permissions. In: Proceedings of the USENIX Conf. on Web Application Development, WebApps (2011)
4. Felt, A.P., Chin, E., Hanna, S., Song, D., Wagner, D.: Android permissions demystified. In: The Proc. of the 18th ACM Conference on Computer and Communications Security, CCS 2011, pp. 627–638. ACM (2011)
5. Sanz, B., Santos, I., Laorden, C., Ugarte-Pedrero, X., Bringas, P.: On the automatic categorisation of android applications. In: Proceedings of the 9th IEEE Consumer Communications and Networking Conf. (CCNC) (2012)
6. Sanz, B., Santos, I., Laorden, C., Ugarte-Pedrero, X., Nieves, J., Bringas, P.G., Álvarez Marañón, G.: MAMA: Manifest Analysis for Malware Detection in Android. Cybernetics and Systems-Intelligent Network Security and Survivability, October 1, 2013
7. Wu, D.-J., Mao, C.-H., Wei, T.-E., Lee, H.-M., Wu, K.-P.: Droidmat: android malware detection through manifest and api calls tracing. In: 2012 Seventh Asia JCIS, pp. 62–69. IEEE, August 2012
8. Barrera, D., Kayacik, H.G., van Oorschot, P.C., Somayaji, A.: A methodology for empirical analysis of permission-based security models and its application to android. In: ACM CCS 2010 (2010)
9. Dini, G., Martinelli, F., Saracino, A., Sgandurra, D.: MADAM: a multi-level anomaly detector for android malware. In: Kotenko, I., Skormin, V. (eds.) MMM-ACNS 2012. LNCS, vol. 7531, pp. 240–253. Springer, Heidelberg (2012)
10. Holmes, G., Donkin, A., Witten, I.H.: Weka: a machine learning workbench, pp. 357–361, August 1994
11. Sahs, J., Khan, L.: A machine learning approach to android malware detection. In: 2012 European Intelligence and Security Informatics Conference (2012)
12. Batyuk, L., Herpich, M., Camtepe, S.A., Raddatz, K., Schmidt, A., Albayrak, S.: Using static analysis for automatic assessment and mitigation of unwanted and malicious activities within android applications. In: 6th International Conference Malicious and Unwanted Software (MALWARE 2011) (2011)
13. Nauman, M., Khan, S., Zhang, X.: Apex: extending android permission model and enforcement with userdefined runtime constraints. In: Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, pp. 328–332. ACM (2010)

14. Ongtang, M., McLaughlin, S., Enck, W., McDaniel, P.: Semantically rich application-centric security in android. In: 2009 Annual Computer Security Applications Conference, pp. 340–349. IEEE (2009)
15. Peiravian, N., Zhu, X.: Machine learning for android malware detection using permission and api calls. In: Proc. of IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp. 300–305. IEEE, November 2013
16. Sato, R., Chiba, D., Goto, S.: Detecting android malware by analyzing manifest files. In: Proceedings of the APAN – Network Research Workshop (2013)
17. Khan, S., Nauman, M., Othman, A.T., Musa, S.: How secure is your smartphone: an analysis of smartphone security mechanisms. In: International Conference on Cyber Security, Cyber Warfare & Digital Forensic (2012)
18. Enck, W., Ongtang, M., McDaniel, P.: On lightweight mobile phone application certification. In: Proc. of the 16th ACM Conference on CCS, pp. 235–245. ACM, November 2009
19. Wei, X., Gomez, L., Neamtiu, I., Faloutsos, M.: Permission Evolution in the Android Ecosystem
20. Zhou, Y., Wang, Z., Zhou, W., Jiang, X.: Hey, you, get off of my market: detecting malicious apps in official and alternative android markets. In: NDSS 2012 (2012)
21. AndroidSecurityOverview. http://source.android.com/devices/tech/security/index.html
22. Google Android Retrieved. http://developer.android.com/guide/basics/what-is-android.html
23. Google Android source code. http://source.android.com/source/downloading.html
24. Google Inc. Android Developers. Manifest. permission. http://developer.android.com/reference/android/Manifest.permission.html
25. Google Inc. Android Developers. Security and Permissions http://developer.android.com/guide/topics/security/security.html
26. Permission>Android Developer - API Guides – Android Manifest. http://developer.android.com/guide/topics/manifest/permission-element.html

Enhancement of Preventing Application Layer Based on DDOS Attacks by Using Hidden Semi-Markov Model

Ko Ko Oo, Kyaw Zaw Ye, Hein Tun, Kyaw Zin Lin and E.M. Portnov

Abstract In this paper, we discuss about the classification of DDOS, which are widely regarded as a major threat to the Internet. Nature of DDoS attack is difficult to defend against and will continue to be an attractive and effective form of attack. Though many solutions have been proposed, the problem has not been solved yet. The defense approaches can be classified as protection, detection and prevention. Therefore, for information security, it is needed to create own DDoS defense system to solve the DDoS attack problems. The aim of the research work is to protect Distributed Denial of Service attacks.

Keywords Information security · Distributed denial of service attacks • Packet classification algorithm · Intrusion detection system · Intrusion prevention system · Hidden semi-markov model

1 Introduction

The Internet has become increasingly important to current society. It is changing our way of communication, business mode, and even everyday life. Almost all the traditional services such as banking, power, medicine, education and defense are extended to Internet now. Internet usage is growing at an exponential rate as organizations, governments and citizens continue to increase their reliance on this technology [1-3].

Unfortunately, security problems are major obstacles to the further development of the Internet. Distributed Denial of Service (DDoS) attack is one of

---

K.K. Oo · K.Z. Ye(✉) · H. Tun · K.Z. Lin · E.M. Portnov
Ministry of Defence, Nay Pyi Daw, Republic of Union of Myanmar,
and Department of Informatics and Computer Software System,
National Research University of Electronic Technology, Zelenograd, Moscow, Russia
e-mail: Kyaw.Z.Ye@ieee.org, kyawzawye85@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_14

the most common and major threat to the Internet. As a side effect, they frequently create network congestion on the way from a source to the target, thus disrupting normal Internet operation. The number of DDoS attack has been alarmingly increasing for the last few years [4, 5].

2 Ease of Use

A web application can be attacked in many ways such as SQL Injection, XSS, DDoS attacks, etc. Among them DDoS attacks are the most difficult to detect and prevent. DDoS attacks occur at two levels such as Network Layer and Application Layer. But Application layer DDoS attacks are very difficult to detect [6] and prevent threads as the protocols used in DDoS attacks are HTTP, DNS, SIP, etc. which can never be blocked by any firewall or IDS/IPS and the pattern is almost similar to normal traffic.

To realize effective Application Layer based attacks to the servers and the better understanding of the DDoS problem for preventing.

- To classify App-DDoS behaviors on network traffic and to distinguish the abnormal HTTP request from the normal ones.
- To evaluate the attack using Hidden semi Markov Model (HsMM) more effectively.
- To enhance in preventing application layer based on DDoS attacks by HsMM.

3 How to Secure System

Protection Approaches

Various steps can be taken to prevent the DDOS attacks to measure. These are called Protective Measures. There are many of them as follows:

(a) Source Validation.
(b) Proof of Work.
(c) Resource Allocation and
(d) Hiding.

Detection Approaches

Among the existing attacks types, the most simple and effective way of App-DDoS attacks is utilizing the HTTP GET to launch attack by requesting home page of the victim website repeatedly. In this paper our detecting schemes consider the App-DDoS attacks as anomaly browsing behavior.

Web browsing behavior can be detected using user request sequences. Usually browsing behavior can describe by using three factors 1) HTTP request rate 2) page viewing time 3) requested sequences. Second way, the user jump from one page to another by typing URLs in address bar, selecting from the favorites of the browser or using navigation tools. Browsing behavior of the user can be calculated by just considering viewing time of the user [7, 8, and 9].

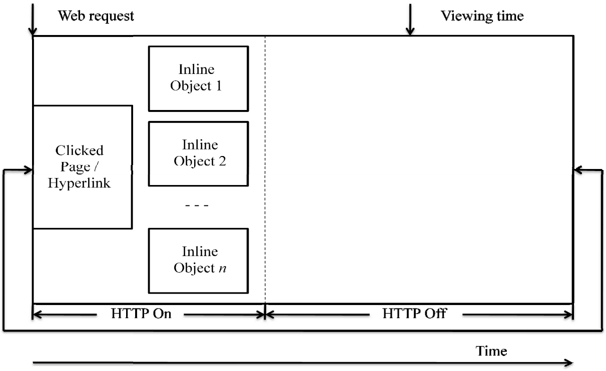


Fig. 1 Diagram of Web Browsing Behavior.

As shown in fig (1), this will reflect the actual time for a page taken by user after requesting a page.

HTTP ON – Requested time of current page.
HTTP OFF – Requested time of next page.

Viewing Time for Current Page = (HTTP OFF – HTTP ON)

Internet proxies and cashes may responds to users HTTP request and thus some of the user's request may not arrive at the web server. Because of this we have to maintain different logs of user's requests for the same page being accessed by different users at different time or different way.

Prevention Approaches

Attack prevention aims to fix security holes, such as insecure protocols, weak authentication schemes and vulnerable computer systems, which can be used as stepping stones to launch a DDoS attack. Attack detection aims to detect DDoS attacks in the process of attacking and characterization helps to distinguish attack traffic from legitimate traffic. Trace back aims to locate the attack sources regardless of the spoofed source IP addresses in either process of attack (active) or after the attack (passive).

4 Roposed System

App-DDoS Prevention System

In this section, the proposed application DDoS attacks prevention system architecture will be described.

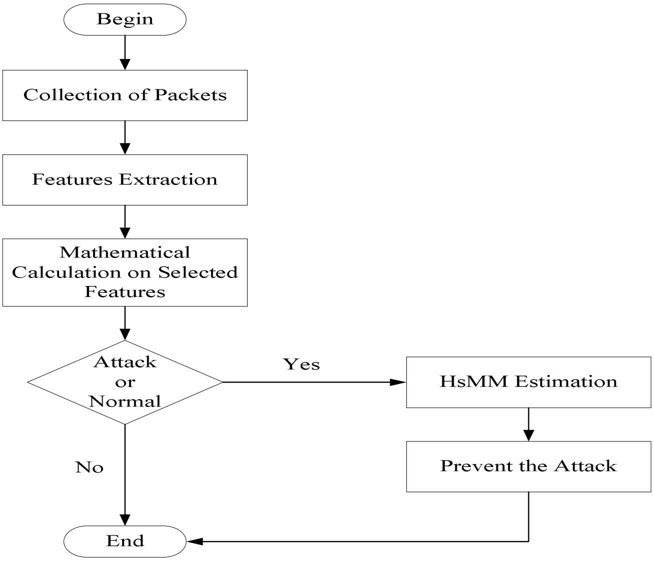

Begin
Collection of Packets
Features Extraction
Mathematical Calculation on Selected Features
Attack or Normal
Yes
HsMM Estimation
Prevent the Attack
No
End

Fig. 2 Flowchart of System Design.

The overall procedure of this prevention architecture is illustrated in fig (2). The scheme is divided into five phases:

(a) Collection of packets.
(b) Features Extraction.
(c) Packet Classification Algorithm.
(d) Estimate Using HsMM and
(e) App-DDoS Attack Prevention.

Collection of Packets

The proposed system collects the incoming packets in every one second. For collecting packets in one second, it does not impact to the classification accuracy. The use of longer flow timeouts is possible; however, there is longer classification response.

Features Extraction

Feature extraction calculates the selected features for captured packets. These features are proposed by observing the characteristics of App-DDoS attack packets. These features can be used to recognize and classify incoming attack packets. The proposed features explain as follows:

(a) Number of packets: Total number of packets from source IP to destination IP. During an attack, the attacker sends a large number of packets to the victim system.

(b) Number of bytes: Total number of bytes sent from source IP to destination IP. It increases when launching DDoS attack.

(c) Average packet size: The ratio of number of bytes to number of packets. It increases in attack time.

(d) Packet rate: Rate of packet per second. For packet rate calculation:

Packet rate per second = $n_p \times \frac{1}{(t_e - t_s)}$, where

n_p = number of packets
t_s= start packet sent time
t_e = end packet sent time

(e) Byte rate: Rate of packets byte per second.

For byte rate calculation:

Byte rate per second = $n_b \times \frac{1}{(t_e - t_s)}$, where

n_b = total number of bytes
t_s = start packet sent time
t_e = end packet sent time

(f) Time-interval variance: The attacker sends attack packets while launching DDoS attack, so time interval variance will be closer to zero.

For time-interval variance calculation:

(1) First, calculating the means:

$$\bar{t} = \frac{\sum t_i}{i} \qquad , i = 1, 2, 3,\ldots$$

(2) Second, squaring deviation of the means

$$t_c^2 = \frac{\sum (t_n - \bar{t})^2}{n}$$, c = collection number (c1, c2, c3, …),

n=no of packets

(3) Third, final calculation of time-interval-variance:

$$t_c = \sqrt{\frac{\sum (t_n - \bar{t})^2}{n}}$$

(g) Packet-size variance: Although normal packets have different packet sizes, attack packet size are the same. So packet-size-variance will be closer to zero.

For packet size calculation:

(1) First, calculating the means:

$$\overline{p} = \frac{\sum p_i}{i}, \qquad i = 1,2,3,\ldots, \text{ where}$$

(2) Second, squaring deviation of the means:

$$p_c^2 = \frac{\sum (p_n - \overline{p})^2}{n} \quad ,c = \text{collection number } (c1, c2, c3,\ldots),$$

n=no of packets

(3)Third, final calculation of packet-size-variance:

$$p_c = \sqrt{\frac{\sum (p_n - \overline{p})^2}{n}}$$

Packet Classification Algorithm

The algorithm checks the field values of the packet. The flowchart in classifies attacks by the field values of the packets.

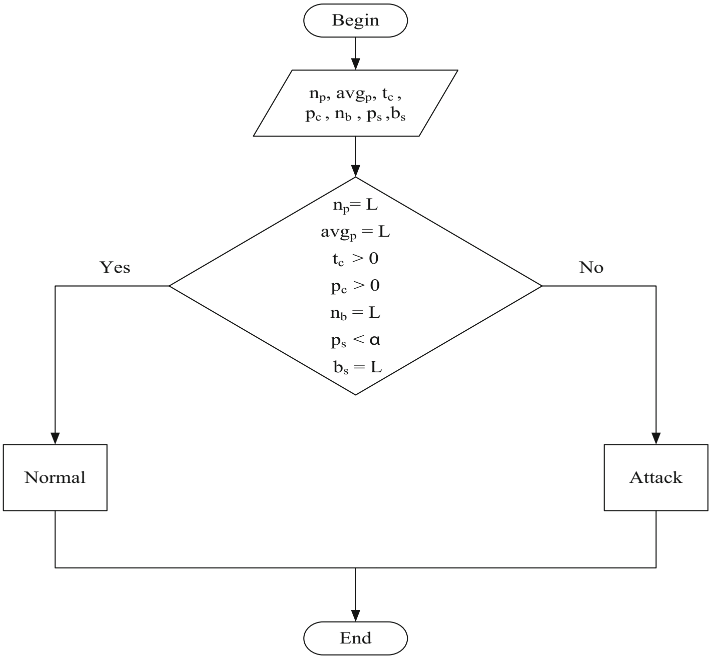

Begin
n_p, avg_p, t_c, p_c, n_b, p_s, b_s
$n_p = L$
$avg_p = L$
$t_c > 0$
$p_c > 0$
$n_b = L$
$p_s < \alpha$
$b_s = L$
Yes
No
Normal
Attack
End

Fig. 3 Packet Classification Algorithm

Table 1 Field Values of Packets for Features Extraction.

Number of packets	**average packet**	**Time - interval-.**	**Packet size-variance**	**Number of bytes**	**Packet rate per**	**Byte rate per second**	**class**
L	L	>0	>0	L	< α	L	Normal
H	H	<0	<0	H	>λ	H	Attack

α = minimum packet rate, λ = maximum packet rate, L = Low, H = High

Estimate Using hsmm

- It is possible that one field value of the incoming packet may miss with the field value of the algorithm.
- For example, in normal case, packet size variance value is not (>0) although other field values match to the algorithm.
- If this case occurs, it cannot be sure that is normal or attack.
- When this case occurs, the proposed system identifies that this case is "Other" case which is not normal or attacks.
- To identify "Other" case as normal or attack, use the HsMM to estimate normal or attack.

An HsMM can be described as

$\Lambda = (H, \Pi, A, P)$, where

(a) H is a discrete set of hidden state H= {1, 2….N}.
(b) Π is the probability distribution for the initial state
(c) A is the state transition matrix
(d) P is the duration Matrix

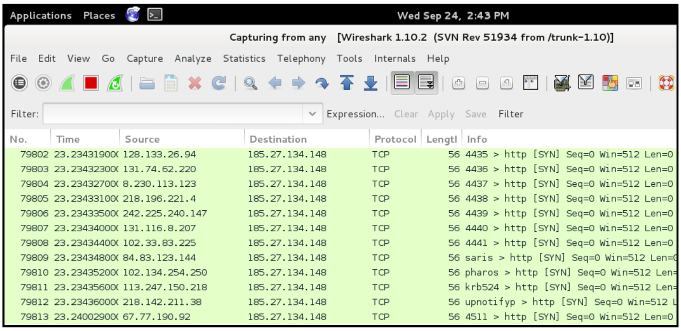
App-DDoS Attack Prevention

Finally, the proposed algorithm classifies and identifies what types of attacks and can be able to prevent App-DDoS.

5 Research Experimentaion

Perform DDoS Attack with Hping Command

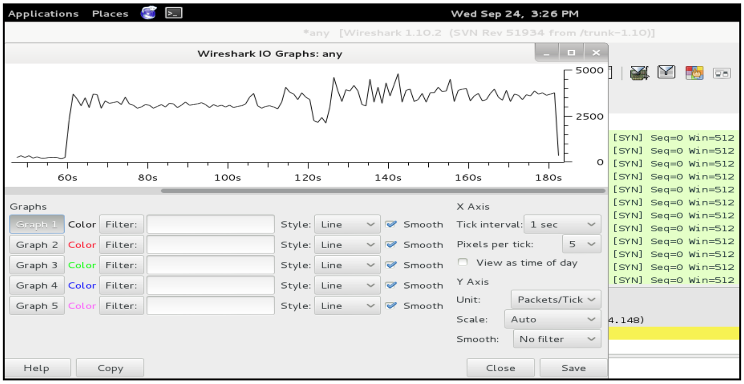
In computing, a Distributed Denial of Service (DDoS) attack is an attempt to make a machine or network resource unavailable to its intended users [10-13]. I have used Hping3 in my study. Using the tool Hping3, we were able to perform SYN flood and HTTP flood DDoS attacks. Hping3 used random source addresses and the attack was performed successfully. Hping3 command for SYN flood attack: hping3 -i u1 -S -p 80 --rand-source 185.27.134.148. Here are also the snapshots captured by traffic capturing tool Wireshark.


Applications
Places
Wed Sep 24, 2:43 PM
Capturing from any [Wireshark 1.10.2 (SVN Rev 51934 from /trunk-1.10)]
File
Edit
View
Go
Capture
Analyze
Statistics
Telephony
Tools
Internals
Help
Filter:
Expression...
Clear
Apply
Save
Filter

```
No.   Time         Source           Destination     Protocol Lengtl Info
79802 23.23431900( 128.133.26.94    185.27.134.148  TCP      56 4435 > http [SYN] Seq=0 Win=512 Len=0
79803 23.23432300( 131.74.62.220    185.27.134.148  TCP      56 4436 > http [SYN] Seq=0 Win=512 Len=0
79804 23.23432700( 8.230.113.123    185.27.134.148  TCP      56 4437 > http [SYN] Seq=0 Win=512 Len=0
79805 23.23433100( 218.196.221.4    185.27.134.148  TCP      56 4438 > http [SYN] Seq=0 Win=512 Len=0
79806 23.23433500( 242.225.240.147  185.27.134.148  TCP      56 4439 > http [SYN] Seq=0 Win=512 Len=0
79807 23.23434000( 131.116.8.207    185.27.134.148  TCP      56 4440 > http [SYN] Seq=0 Win=512 Len=0
79808 23.23434400( 102.33.83.225    185.27.134.148  TCP      56 4441 > http [SYN] Seq=0 Win=512 Len=0
79809 23.23434800( 84.83.123.144    185.27.134.148  TCP      56 saris > http [SYN] Seq=0 Win=512 Len=(
79810 23.23435200( 102.134.254.250  185.27.134.148  TCP      56 pharos > http [SYN] Seq=0 Win=512 Len=
79811 23.23435600( 113.247.150.218  185.27.134.148  TCP      56 krb524 > http [SYN] Seq=0 Win=512 Len=
79812 23.23436000( 218.142.211.38   185.27.134.148  TCP      56 upnotifyp > http [SYN] Seq=0 Win=512 L
79813 23.24002900( 67.77.190.92     185.27.134.148  TCP      56 4511 > http [SYN] Seq=0 Win=512 Len=0
```

Fig. 4 SYN Flood Attack on 185.27.134.148 (Random source addresses).

In this case, the source is randomized by the hping (using --rand-source) command. The default protocol while using hping DDoS is NBNS protocol. Given below is the IO graph showing the duration for which the attack was performed. As shown in the Fig (4) and (5) that the attack was performed for less than 5 secs and hping3 sent more than 1 million packets in this small interval of time.


Applications
Places
Wed Sep 24, 3:26 PM
*any [Wireshark 1.10.2 (SVN Rev 51934 from /trunk-1.10)]
Wireshark IO Graphs: any
5000
2500
0
60s
80s
100s
120s
140s
160s
180s
Graphs
Graph 1
Graph 2
Graph 3
Graph 4
Graph 5
Color
Filter:
X Axis
Y Axis
4.148)
Help
Copy
Close
Save

```
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
[SYN] Seq=0 Win=512
```

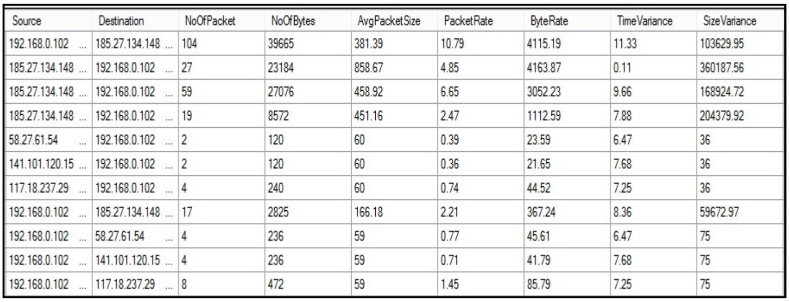
Style: Line ☑ Smooth
Style: Line ☑ Smooth
Style: Line ☑ Smooth
Style: Line ☑ Smooth
Style: Line ☑ Smooth
Tick interval: 1 sec
Pixels per tick: 5
☐ View as time of day
Unit: Packets/Tick
Scale: Auto
Smooth: No filter

Fig. 5 IO Graph of SYN Flood.

Experimental Evaluation

All machine learning experiments were done by the Microsoft Virtual Studio in C#, version 2010. The input of log dataset is captured with Wireshark and then we perform features extraction and packet classification operations are applied.

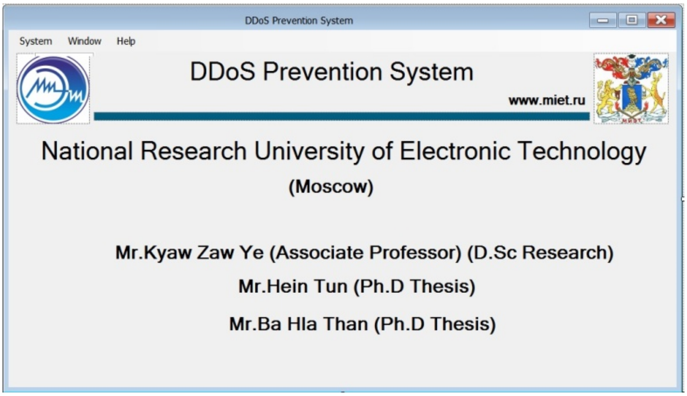
Then the packets were classified by hand into 7 application classes as the base for experimentations, namely No of packet, No of Bytes, Avg Packet Size, Packet size, Byte Rate, Time Variance and Size variance. The following table presents the mathematical calculation for features extraction.



Source	Destination	NoOfPacket	NoOfBytes	AvgPacketSize	PacketRate	ByteRate	TimeVariance	SizeVariance
192.168.0.102 ...	185.27.134.148 ...	104	39665	381.39	10.79	4115.19	11.33	103629.95
185.27.134.148 ...	192.168.0.102 ...	27	23184	858.67	4.85	4163.87	0.11	360187.56
185.27.134.148 ...	192.168.0.102 ...	59	27076	458.92	6.65	3052.23	9.66	168924.72
185.27.134.148 ...	192.168.0.102 ...	19	8572	451.16	2.47	1112.59	7.88	204379.92
58.27.61.54 ...	192.168.0.102 ...	2	120	60	0.39	23.59	6.47	36
141.101.120.15 ...	192.168.0.102 ...	2	120	60	0.36	21.65	7.68	36
117.18.237.29 ...	192.168.0.102 ...	4	240	60	0.74	44.52	7.25	36
192.168.0.102 ...	185.27.134.148 ...	17	2825	166.18	2.21	367.24	8.36	59672.97
192.168.0.102 ...	58.27.61.54 ...	4	236	59	0.77	45.61	6.47	75
192.168.0.102 ...	141.101.120.15 ...	4	236	59	0.71	41.79	7.68	75
192.168.0.102 ...	117.18.237.29 ...	8	472	59	1.45	85.79	7.25	75

Fig. 6 Mathematical Calculation for Features Extraction.

The classification process focuses on features extraction. For the classification purpose, a set of packet statistical features are extracted discretized to represent packet flows. The Fig (7) shows the App-DDoS prevention system window.



DDoS Prevention System

www.miet.ru

National Research University of Electronic Technology

(Moscow)

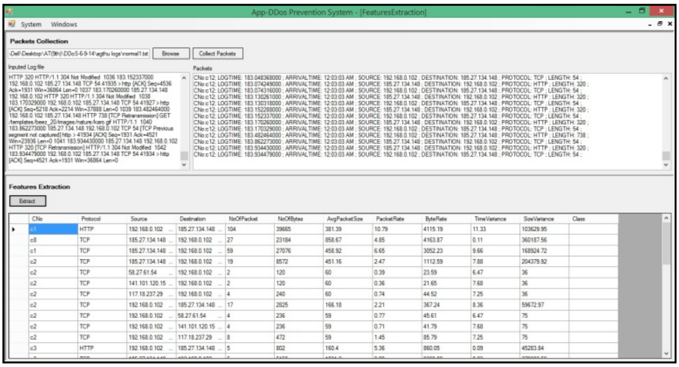
Mr.Kyaw Zaw Ye (Associate Professor) (D.Sc Research)

Mr.Hein Tun (Ph.D Thesis)

Mr.Ba Hla Than (Ph.D Thesis)

Fig. 7 App-DDoS Prevention System.

This window will appear first when we run the program and selecting system menu for packet classification. The Fig (8) shows the features extraction window. In this the packets are extracted as No of packet, No of Bytes, Average Packet Size, Packet size, Byte Rate, Time Variance invariance.



\Dell\Desktop\AT(9th)\DDoS-6-9-14\agthu logs\normal1.txt

Inputed Log file

```
HTTP 320 HTTP/1.1 304 Not Modified  1036 183.152337000
192.168.0.102 185.27.134.148 TCP 54 41935 > http [ACK] Seq=4536
Ack=1931 Win=36864 Len=0 1037 183.170260000 185.27.134.148
192.168.0.102 HTTP 320 HTTP/1.1 304 Not Modified  1038
183.170329000 192.168.0.102 185.27.134.148 TCP 54 41927 > http
[ACK] Seq=5218 Ack=2214 Win=37888 Len=0 1039 183.482464000
192.168.0.102 185.27.134.148 HTTP 738 [TCP Retransmission] GET
/templates/beez_20/images/nature/karo.gif HTTP/1.1  1040
183.862273000 185.27.134.148 192.168.0.102 TCP 54 [TCP Previous
segment not captured] http > 41934 [ACK] Seq=1931 Ack=4521
Win=23936 Len=0 1041 183.934430000 185.27.134.148 192.168.0.102
HTTP 320 [TCP Retransmission] HTTP/1.1 304 Not Modified  1042
183.934479000 192.168.0.102 185.27.134.148 TCP 54 41934 > http
[ACK] Seq=4521 Ack=1931 Win=36864 Len=0
```

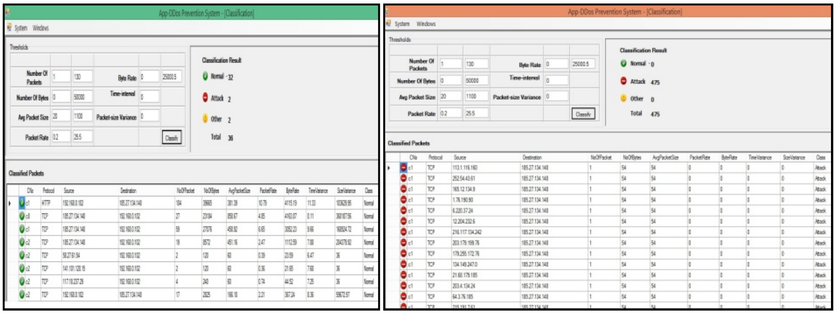
Packets

```
CNo:c12: LOGTIME: 183.048368000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 192.168.0.102 ; DESTINATION: 185.27.134.148 ; PROTOCOL: TCP ; LENGTH: 54 ;
CNo:c12: LOGTIME: 183.074249000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 185.27.134.148 ; DESTINATION: 192.168.0.102 ; PROTOCOL: HTTP ; LENGTH: 320 ;
CNo:c12: LOGTIME: 183.074316000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 192.168.0.102 ; DESTINATION: 185.27.134.148 ; PROTOCOL: TCP ; LENGTH: 54 ;
CNo:c12: LOGTIME: 183.130261000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 185.27.134.148 ; DESTINATION: 192.168.0.102 ; PROTOCOL: HTTP ; LENGTH: 320 ;
CNo:c12: LOGTIME: 183.130318000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 192.168.0.102 ; DESTINATION: 185.27.134.148 ; PROTOCOL: TCP ; LENGTH: 54 ;
CNo:c12: LOGTIME: 183.152288000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 185.27.134.148 ; DESTINATION: 192.168.0.102 ; PROTOCOL: HTTP ; LENGTH: 320 ;
CNo:c12: LOGTIME: 183.152337000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 192.168.0.102 ; DESTINATION: 185.27.134.148 ; PROTOCOL: TCP ; LENGTH: 54 ;
CNo:c12: LOGTIME: 183.170260000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 185.27.134.148 ; DESTINATION: 192.168.0.102 ; PROTOCOL: HTTP ; LENGTH: 320 ;
CNo:c12: LOGTIME: 183.170329000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 192.168.0.102 ; DESTINATION: 185.27.134.148 ; PROTOCOL: TCP ; LENGTH: 54 ;
CNo:c12: LOGTIME: 183.482464000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 192.168.0.102 ; DESTINATION: 185.27.134.148 ; PROTOCOL: HTTP ; LENGTH: 738 ;
CNo:c12: LOGTIME: 183.862273000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 185.27.134.148 ; DESTINATION: 192.168.0.102 ; PROTOCOL: TCP ; LENGTH: 54 ;
CNo:c12: LOGTIME: 183.934430000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 185.27.134.148 ; DESTINATION: 192.168.0.102 ; PROTOCOL: HTTP ; LENGTH: 320 ;
CNo:c12: LOGTIME: 183.934479000 ; ARRIVALTIME: 12:03:03 AM ; SOURCE: 192.168.0.102 ; DESTINATION: 185.27.134.148 ; PROTOCOL: TCP ; LENGTH: 54 ;
```

CNo	Protocol	Source	Destination	NoOfPacket	NoOfBytes	AvgPacketSize	PacketRate	ByteRate	TimeVariance	SizeVariance	Class
c1	HTTP	192.168.0.102 ...	185.27.134.148 ...	104	39665	381.39	10.79	4115.19	11.33	103629.95	
c8	TCP	185.27.134.148 ...	192.168.0.102 ...	27	23184	858.67	4.85	4163.87	0.11	360187.56	
c1	TCP	185.27.134.148 ...	192.168.0.102 ...	59	27076	458.92	6.65	3052.23	9.66	168924.72	
c2	TCP	185.27.134.148 ...	192.168.0.102 ...	19	8572	451.16	2.47	1112.59	7.88	204379.92	
c2	TCP	58.27.61.54 ...	192.168.0.102 ...	2	120	60	0.39	23.59	6.47	36	
c2	TCP	141.101.120.15 ...	192.168.0.102 ...	2	120	60	0.36	21.65	7.68	36	
c2	TCP	117.18.237.29 ...	192.168.0.102 ...	4	240	60	0.74	44.52	7.25	36	
c2	TCP	192.168.0.102 ...	185.27.134.148 ...	17	2825	166.18	2.21	367.24	8.36	59672.97	
c2	TCP	192.168.0.102 ...	58.27.61.54 ...	4	236	59	0.77	45.61	6.47	75	
c2	TCP	192.168.0.102 ...	141.101.120.15 ...	4	236	59	0.71	41.79	7.68	75	
c2	TCP	192.168.0.102 ...	117.18.237.29 ...	8	472	59	1.45	85.79	7.25	75	
c3	HTTP	192.168.0.102 ...	185.27.134.148 ...	5	802	160.4	5.36	860.05	0.09	45283.84	

Fig. 8 App-DDoS System (Features Extraction)

The following Fig (9) shows the classification windows. In the Figures described the attack or normal of received packets. In this the packets are classified as Normal packets, Attack packets and others.



Number Of Packets	1	130	Byte Rate	0	25000.5
Number Of Bytes	0	50000	Time-interval	0	
Avg Packet Size	20	1100	Packet-size Variance	0	
Packet Rate	0.2	25.5			

Normal - 32
Attack 2
Other 2
Total 36

CNo	Protocol	Source	Destination	NoOfPacket	NoOfBytes	AvgPacketSize	PacketRate	ByteRate	TimeVariance	SizeVariance	Class
c1	HTTP	192.168.0.102	185.27.134.148	104	39665	381.39	10.79	4115.19	11.33	103629.95	Normal
c8	TCP	185.27.134.148	192.168.0.102	27	23184	858.67	4.85	4163.87	0.11	360187.56	Normal
c1	TCP	185.27.134.148	192.168.0.102	59	27076	458.92	6.65	3052.23	9.66	168924.72	Normal
c2	TCP	185.27.134.148	192.168.0.102	19	8572	451.16	2.47	1112.59	7.88	204379.92	Normal
c2	TCP	58.27.61.54	192.168.0.102	2	120	60	0.39	23.59	6.47	36	Normal
c2	TCP	141.101.120.15	192.168.0.102	2	120	60	0.36	21.65	7.68	36	Normal
c2	TCP	117.18.237.29	192.168.0.102	4	240	60	0.74	44.52	7.25	36	Normal
c2	TCP	192.168.0.102	185.27.134.148	17	2825	166.18	2.21	367.24	8.36	59672.97	Normal

Number Of Packets	1	130	Byte Rate	0	25000.5
Number Of Bytes	0	50000	Time-interval	0	
Avg Packet Size	20	1100	Packet-size Variance	0	
Packet Rate	0.2	25.5			

Normal - 0
Attack 475
Other 0
Total 475

CNo	Protocol	Source	Destination	NoOfPacket	NoOfBytes	AvgPacketSize	PacketRate	ByteRate	TimeVariance	SizeVariance	Class
c1	TCP	113.1.116.160	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	252.54.43.61	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	165.12.134.9	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	1.76.190.90	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	6.220.37.24	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	12.204.232.6	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	216.117.134.242	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	203.179.199.76	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	179.255.172.76	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	134.149.247.0	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	21.68.179.185	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	203.4.134.24	185.27.134.148	1	54	54	0	0	0	0	Attack
c1	TCP	64.3.76.185	185.27.134.148	1	54	54	0	0	0	0	Attack

Fig. 9 App-DDoS Prevention System (Classification) with Normal Dataset and Hping Dataset.

Many studies on DDoS attack detection have been carried out; however, they focus only on the change in network traffic. My research works based dataset are suitable for the detection; however, they do not ensure real-time transfer of packets. Hence, my research work can classify the current network status well to detect DDoS attacks early.

6 Conclusion

The focus on this research work on detecting App-DDoS attacks. HsMM describe attack type of App-DDoS by training the model from set of request sequences made by lot of normal packets. In the proposed system, App-DDoS attacks packets are studied from collection of data set and extract the features to analyze and classify App-DDoS attack. The proposed algorithm and packet classification algorithm will be efficient and suitable for classifying flooding attacks and vulnerability attacks.

Acknowledgements The research work was supported by National Research University of Electronic Technology and Laboratory of Information management systems and complexes Moscow, Russian Federation.

References

1. Douligeris, C., Mitrokotsa, A.: DDoS Attacks and Defense Mechanisms: Classification and State of the Art. Computer Journal of Networks **44**(5), 643–666 (2004)
2. Babu, G.P., Jayavani, V., Mohan Rao, C.P.V.N.J.: Anomaly Detection On User Browsing Behaviors Using Hidden Semi-Markov Model. IJCSIT. ISSN:0975-9646
3. Bächer, P., et al.: Know your enemy: Tracking botnets. The Honeynet Project and Research Alliance, Tech. Rep. (2005)
4. Garcia-Teodoro, P., Diaz-Verdejo, J., MaciaFernandez, G., Vazquez, E.: Anomaly-Based Network Intrusion Detection: Techniques. Systems and Challenges. Computers and Security **28**(1–2), 18–28 (2009)
5. Radware Ltd. DDoS survival handbook (2013)
6. Lee, S., Kim, G., Kim, S.: Sequence-OrderIndependent Network Profiling for Detecting Application Layer DDos Attacks. EURASIP J. Wireless Comm. and Networking **2011**(1), 50 (2011)
7. Yu, S., Zhou, W., Doss, R., Jia, W.: Traceback of DDos Attacks Using Entropy Variations. IEEE Trans. Parallel and Distributed Systems **22**(3), pp. 412–442
8. Peng, T., Leckie, C., Ramamohanarao, K.: Survey of Network Based Defense Mechanisms Countering the Dos and DDos Problems. ACM Computing Surveys **39**(1), 3 (May 2007) (March 2011)
9. Yi, X., Shunzheng, Y.: A dynamic anomaly detection model for web user behavior based on HsMM. In: Proc. 10th Int. Conf. Computer Supported Cooperative Work in Design (CSCWD 2006), Nanjing, China, vol. 2, pp. 811–816, May 2006
10. Xie, Y., Yu, S.: A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors. IEEE/ACM Trans. Networking **17**(1), 54–65 (2009)

Heuristic Rules for Attack Detection Charged by NSL KDD Dataset

Khaing Shwe Wutyi and Mie Mie Su Thwin

Abstract With the rapidly growing and wide spread use of computer networks, the number of new threats has grown extensively. Automated rule induction procedures for detecting these threats, like machine learning and statistical techniques result in rules that lack generalization and maintainability. In this paper, we focus on detailed study of different types of attacks using NSL KDD dataset by manually developing rules through incorporation of attack signatures. It results in meaningful but weak rules as it is difficult to define thresholds. This paper utilizes a hybrid procedure for developing rules by combining expert knowledge with automated techniques to improve readability, comprehensibility, and maintainability of rules. Through the proposed rule-formation technique, heuristic rules were developed for different attack types included in NSL KDD dataset. Empirical results show that high detection rates with low false alarms are observed for different attack types in the dataset. The utilized techniques also highlighted a mislabeling problem in the NSL KDD dataset for the R2L and U2R attacks considered.

Keywords Hidden-Semi Markov Model (HSMM) · NSL KDD dataset · Heuristic rule · Denial of Service (DoS) · Remote to Local (R2L) · User to Root (U2R) · Probing · NSL KDD Dataset · Detection probability · False alarm rate

1 Introduction

Heuristic rules have been commonly used to detect different types of attacks included in NSL KDD dataset [1], [2], [3]. Different attack categories offer the most diverse set of attacks in terms of attack execution, implementation and dynamics: attacks in NSL KDD dataset differ vastly in terms of signatures and the host against which they are executed. The divers knowledge required to detect

---

K.S. Wutyi(✉) · M.M.S. Thwin
University of Computer Studies Mandalay (UCSM), Mandalay, Myanmar
e-mail: {wutyi.pan,miemiesuthwinster}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_15

attacks inspired many expert systems in literature including P-BEST [4], EMERALD [5] and RIPPER [6].

The advantage of using heuristic rules (data or signature based) is that they can detect attacks in real-time as they typically require very little processing time. There are a number of limitations of heuristic rules as well. It is not a trivial task to develop precise rules, at least for certain applications, because human expert knowledge might not readily lend itself to an effective formulation in the form of an IF-THEN rule template. It is also a challenging task to write rules that are robust enough to work in unknown environments where there is significant noise as the rules might not offer adequate generalization capabilities.

2 Literature Review

The recent literature presents a set of significant efforts to detect R2L attacks using heuristic rules in networked computing environments: specifically the Defense Advanced Research Project Agency (DARPA) and Knowledge Discovery in Databases (KDD) datasets were utilized for most of these studies. In 1998, DARPA funded an "Intrusion Detection Evaluation Program (IDEP)" in Lincoln Laboratory at the Massachusetts Institute of Technology [7]. Three kinds of data was collected: transmission control protocol (TCP) packets using the "tcpdump" utility, basic security module (BSM) audit records using the Sun Solaris BSM utility, and system file dumps. Stolfo and Lee [3], one of the participants in DARPA 1998 program, used TCP packets to build the KDD dataset, which consisted of records based on individual TCP sessions. Each record had 41 features and the method used to derive these features is discussed in [8]. Data-mining techniques were utilized to generate features using the TCP packets for different connections. The KDD dataset is accessible through the UCI KDD archive [9].

Agarwal and Joshi [1] proposed a two-stage general-to-specific framework for learning a rule-based model (PNrule) to learn classifier models on a data set that has widely different class distributions in the training data set.

Levin [2] used Kernel Miner tool on the KDD data set. Kernel Miner is data-mining tool for classification of data and prediction of new cases using automatically generated decision trees. Using this tool and the KDD data set, Levin created a set of locally optimal decision trees (called the decision forest) from which optimal subset of trees (called the sub-forest) was selected for predicting new cases. Levin used only 10% of the KDD training data set randomly sampled from the entire training data set

Yeung and Chow [10] proposed a novelty detection approach using non-parametric density estimation based on Parzen-window estimators with Gaussian kernels to build an intrusion detection system using normal data only. This novelty detection approach was employed to detect attack categories in the KDD data set. 30,000 randomly sampled normal records from the KDD training data set were used as training data to estimate the density of the model.

Lee and Stolfo [11] used data mining techniques to collect KDD features from DARPA 1998 dataset. The proposed model could detect only 20% of R2L attacks with a false alarm rate of 0.01 for the DARPA 1998 testing dataset.

The results of DARPA 1998 IDS evaluation program [12] indicate that the best model proposed was EMERALD that could detect only 35% of R2L attacks in the testing dataset. KDD dataset was used in the UCI KDD 1999 competition. The aim of the competition was to develop intrusion detection system models to address attack categories Probe, DoS, U2R, and R2L. Results of the KDD Cup 1999 competition [13] indicate that the winner was able to correctly classify only 7.82% of R2L records present in the KDD testing dataset.

This paper shall propose heuristic rules that will be created by analyzing the signatures of different attacks in the NSL KDD Dataset. The rule formulation technique will use signature analysis to define a set of features and use automated techniques to assign various threshold values needed in the formulation of heuristic rules. A comprehensive and detailed understanding of different attacks dynamics, mechanisms, and signatures is likely to help develop heuristic rules that will not only potentially offer high probability of detection but also a low false alarm rate. This paper shall analyze various attacks by combining both KDD training and testing datasets and propose heuristic rules to improve detection rate and reduce false alarms. The advantage of using signature-based heuristic rules is that they are not affected by noise in the training dataset. Hence well-formed, signature-based heuristic rules are expected to detect the attacks with low false and missed alarm rates if the NSL KDD feature set can potentially provide adequate observables of attack dynamics.

3 NSL KDD Intrusion Detection Dataset

NSL KDD intrusion detection datasets [7] which are based on DARPA ’98 dataset provides labelled data for researcher working in the field of intrusion detection and is the only labelled dataset publicly available. Normal connections are created to profile that expected in a military network and attacks fall into one of the four categories:

- **Denial of Service (DoS):** Attacker tries to prevent legitimate users from using a service.
- **Remote to Local (R2L):** Attacker does not have an account on the victim machine, hence tries to gain access.
- **User to Root (U2R):** Attacker has local access to the victim machine and tries to gain super user privileges.
- **Probe:** Attacker tries to gain information about the target host. There are 41 features for each connection, which are detailed in Table I. Specifically, “A connection is a sequence of TCP packets starting and ending at some well-defined times, between which data flows from a source IP address to a target IP address under some well-defined protocol”. Features are grouped into four categories:

- **Basic Features:** Basic features can be derived from packet headers without inspecting the payload.
- **Content Features:** Domain knowledge is used to access the payload of the original TCP packets. This includes features such as number of failed login attempts.
- **Time-based Traffic Features:** These features are designed to capture properties that mature over a 2 second temporal window. One example of such a feature would be the number of connections to the same host over the 2 second interval.
- **Host-based Traffic Features:** Utilize a historical window estimated over the number of connections instead of time. Host-based features are designed to access attacks, which span intervals longer than 2 seconds.

Table 1 Attack types with their corresponding categories

Normal	Probing	DoS	R2L	U2R
	Nmap	5. Land	11. Spy	19.Buffer_over flow
	Portsweep	6. Pod	12. Phf	20. Rootkit
	Ipsweep	7. Teardrop	13. Multihop	21. Loadmodule
	Satan	8. Back	14. ftp_write	22. Perl
		9. Neptune	15. Imap	
		10. Smurf	16. Warezmaster	
			17. warezclient	
			18. Guess_Passwd	

4 Formulation of Heuristic Rules for R2L Attacks

In this section, signatures of various NSL KDD dataset attacks will be analyzed. The aim will be to extract relevant features from signatures that must be selected to conclusively observe the attack in a networked environment. These rules will directly map the attack signatures. Since these rules will be tested on the NSL KDD testing data set, an attempt shall be made to formulate rules in terms of the NSL KDD features. Various thresholds (in premises of rules) will be set as they are observed in the NSL KDD datasets. The proposed rules shall be discussed with respect to both attack signatures and the feature sets to facilitate understanding of the mapping between the two.

Thresholds for different rules created were obtained as follows. First the NSL KDD training and testing datasets were merged. For a given attack under consideration in this paper, each record in the merged NSL KDD dataset was labeled as ether belonging to or not belonging to that attack. Expert knowledge was then applied on this relabeled dataset to obtain the rule set for the same attack. These rule sets were consulted whenever needed to define thresholds for features. This technique can easily be extended to other attacks as well.

4.1 Spy Attack

A remote to local (R2L) attack is a class of attacks where an attacker sends packets to a machine over network, then exploits the machine's vulnerability to illegally gain local access to a machine. It occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine. In order to verify spy attack, we define the rule according to the relevant feature set and value of each feature set.

```
If Duration = 299 Λ 337 Λ
Protocol Type = tcp Λ
Service = telnet Λ
Flag = SF               Spy Attack …. 4.1
```

4.2 Phf Attack

The goal of these attacks is to access some resources that your machine should not provide the attacker. For example, a host might be a web server, and should provide anyone with requested web pages. However, that host should provide command shell access without being sure that the person making such a request is someone who should get it, such as a local administrator. This kind of attack can be from local user who need to abuse administrative privilege or it can also be form remote users. We can identify one rule for detecting the phf attack and their relevant and most dependent feature set.

```
If Duration = 0 Λ 6 Λ 12 Λ
Protocol Type = tcp Λ
Service = http Λ
Flag = SF               then output = Phf Attack …. 4.2
```

4.3 Imap Attack

This is a sub attack of R2L attack type. It is a class of attack where an attacker sends packets to a machine over a network, then exploits machine's vulnerability to illegally gain local access as a user; unauthorized access from a remote machine. We can identify the most relevant feature set for detecting Imap attack. Firstly, we organize one rule according to the values of these feature set.

```
If Duration = 0 Λ 31 Λ 41 Λ
Protocol Type = tcp Λ
Service = imap4 Λ
Flag = SF Λ S1 Λ SH    Then output = Imap Attack …. 4.3
```

4.4 Warezmaster Attack

Warezmaster exploits a system bug associated with a file transfer protocol (FTP) server. This attack takes place when an FTP server has, by mistake, given write permissions to users on the system. Hence any user can login and upload files. During the execution of the attack, the attacker logs on the server using the guest account. The attacker then creates a hidden directory and uploads "warez" (copies of illegal software) onto the server. Other users can then later download these files. One simple and obvious way to prevent this attack is to assign correct permissions to the users on the FTP server. "If during an FTP session, large amount is data is sent from source as compared to destination then warezmaster attack can be concluded."

```
If (Duration > 265) Λ
(Protocol = tcp) Λ
(Service = ftp V ftp_data) Λ
(source_bytes > 265616) Λ
(destination_bytes = 0) Then output = Warezmaster
Attack… Rule 4.4
```

4.5 Warezclient Attack

Warezclient attack can be launched by any legal user during an FTP connection after warezmaster attack has been executed. During warezclient attack, users download the illegal "warez" software that was posted earlier through a successful warezmaster attack. Since this process requires downloading files from the FTP server, attack dynamics project a perfectly legal process. The rule to detect warezclient attack in compliance with the NSL KDD dataset features can be stated as follows:

"If a user, during an FTP session, triggers notably many hot indicators to be set in a small duration of time then the user maybe downloading illegally posted software from the server."

In terms of the KDD features the rule can be defined as:

```
If (Duration < 5) Λ
(Protocol = tcp) Λ
(Service = ftp V ftp_data) Λ
(logged_in = 1 V is_guest_login = 1)
(Hot > 25)     Then output = Warezclient Attack… Rule 4.5
```

5 Nature of Feature Set and Value for Different Attack Types

Depending on the nature of feature set and value, we can define the different attack types. We can also identify prominent, dependent and Neglectable feature set and value for each attack type. By doing so, three cases can be able to identify for each feature set and attack. These cases are normal case, abnormal (anomalous) case and attack case. Each known attack pattern can be compared and decided to whether it is normal case or anomalous case. So the system can provide early warning stage for users because it can predict the changes on the values and features of each incoming dataset.

Within these three cases, we can also differentiate whether it is completely normal or anomalous case. For example, in one attack case, the first three feature sets and values predict that it will be one attack case but suddenly it changes into non-attack case after checking next feature set and value. This kind of cases can happen in normal and anomalous case. We can compare the incoming feature set and value pattern with the known attack and normal pattern of NSL KDD dataset.

We can also guarantee about the results and patterns of our system because NSL KDD dataset is the most standardized dataset for detecting network attack types. In order to find the patterns for each attack types and normal process, firstly we need to identify which are the most prominent, dependent and Neglectable feature set for each attack. For doing so, defining heuristic rules are the most simplest and effective ways for defining patterns for each attack and normal case. By defining heuristic rules, we can identify not only the prominent feature set but also the value of these feature sets. We need to analyze thoroughly about these feature set and values because changes on one value of these feature can affect the patterns of different attacks.

On previous research, most of the researchers use machine learning programs, data mining methods and others statistical model to find out the different attacks in NSL KDD dataset. Previous research didn’t find out the nature of dataset’s feature and value which is very import for examining whether it can be attack or not. Our research aims to find out the relationship and patterns of these features depending on the different attack types which is included in NSL KDD dataset which is a standardized dataset for network attack intrusion dataset.

5.1 Prominent Feature Set and Value

The nature of the prominent feature set and value should be only one for each attack type. However, in some cases, it is very difficult to define the nature of each attack type by only one prominent feature. But sometimes, we can identify easily with only one feature set. In parallel, some cases are same to define different attack pattern.

Table 2 Most Prominent Features and its Value of R2L attacks

Attack Name	Prominent Feature	Value
Spy	Diff_srv_rate, dst_host_srv_count	0, 0
Phf	Rerror_rate, srv_rerror_rate	0 to 1, 0 to 1
Imap	Srv_error_rate, diff_srv_rate	0 to 1, 0 to 1
Warezmaster	Num_rood, num_file_creations, num_shells	0 to 7000, 0 to 43, 0 to 2
Warezclient	Num_access_files, num_outbound_cmds	0 to 9, 0

5.2 Neglectable Feature Set and Value

Finding the Neglectable feature set and its value is also important for defining the patterns of different attack types. Because if we neglect some of the feature and its value but the changes on that neglected features and its value affect the pattern of attack. So it will be difficult to examine the nature of feature set and value changes. In NSL KDD dataset, there are total 41 feature set and its values ranges from binary to numeric types. So we need to find clearly which features can be neglected among 41 features. By doing so, we can select the most simple and easiest way to define normal and anomalous behavior of different attack case.

5.3 Normal Case

For the perfect normal case, there is no deviations from predefining heuristic rules and attack patterns. Although it takes to define these patterns depending on huge amount of dataset but it's worthy. Like the previous anomalous behavior pattern mining, if we know the normal case for both attack and normal process, we can predict the anomalous behavior of the incoming feature set. The following is the perfect normal case for both normal and attack pattern.

Table 3 Perfect Normal Case for normal pattern

Duration	Protocol Type	Service	Flag	Src_bytes	srv_count	Dst_host_srv_rerror_rate	
0	tcp	ftp_data	SF	491	2	0	normal
0	udp	other	SF	146	1	0	normal
0	tcp	http	SF	232	5	0.01	normal
0	tcp	http	SF	199	32	0	normal
0	tcp	http	SF	287	7	0	normal
0	tcp	http	SF	300	9	0	normal
0	tcp	http	SF	233	3	0	normal

5.4 Attack Case

For the attack case, there can be two types. One is perfectly normal attack case and other is anomalous case where we can provide the warning stage. The following is the perfectly normal attack case.

Table 4 Perfect Normal Case for Attack Pattern (Phf – R2l attack)

Duration	Protocol Type	Service	Flag	
0	tcp	http	SF	phf
12	tcp	http	SF	phf
6	tcp	http	SF	phf
0	tcp	http	SF	phf

5.5 Anomalous Case

Any case that is deviated from the predefined, normal and attack pattern, can be defined as anomalous case. But we can divide roughly two cases. One is anomalous case for normal pattern and the other one is anomalous case for attack pattern. The following is the examples of these cases.

Table 5 Anomalous Case for Attack Pattern (Phf – R2L attack)

Duration	Protocol Type	Service	Flag	
0	tcp	http	SF	phf
12	tcp	http	SF	phf
6	tcp	http	SF	phf
0	tcp	http	SF	phf
88892390	tcp	http	SF	Anomalous

As shown in above table, although the feature set or pattern seems to be happen Phf – R2L attack, it changes suddenly into anomalous case. The duration for anomalous case is more than that of previous normal or attack case. So after checking 3 features and its values, the 4th one seems to be different from normal one. So we can conclude that one is abnormal pattern which is seems to be normal or phf attack pattern.

6 Comparative View

Based on this short and selective survey of incremental techniques for anomaly detection, we make the following observations.

6.1 Pros and Cons of the Comparative Methods and Detection Accuracy

The following table is the pros and cons of the comparative methods and detection accuracy comparing to comparative methods. According to this, our system seems to be quite promising then other methods and provide high detection accuracy.

Table 6 Pros and Cons of the methods

Method	Summary		Performance	
	Pros	**Cons**	**Detection Rate (%)**	**False Positive Rate (%)**
SC+ITI based [14]	• Performs well over force assignment problem and class dominance problem in terms of detection rate and false positive rate.	• Performance comparison with other relevant techniques missing. • Not tested in a real-time environment.	92.63	1.80
SVM [15]	• Tested their approach in two possible scenarios: (*i*) learning with limited resources and (*ii*) active learning. They claim that it performs in constant running time during learning as well as prediction.	• Too many input parameters.	-	-
Density-based [16]	• Reduces the cost of re-clustering and profile updating and claims to achieve higher accuracy than ADWICE.	• In parameter level, it does not make any improvement. • Lack of performance results over larger datasets.	-	-
EoHMMs [17]	• On synthetic and real-world HIDS datasets, achieves higher detection accuracy when single best HMM are estimated at each learning stage.	• Not explored the dynamic selection of the best classifiers, decision threshold, and Boolean fusion function during the operations.	-	-
RS-ISVM [18]	• Reduces the training time as well as prediction time. • Performs better than the simple SVM in terms of detection rate and false positive rate	• Performance is poor in case of U2R and R2L attacks. • It needs to reduce the number of parameters during training phase.	89.17	4.9
iECT [19]	• Capable to update the eigenvectors and eigenvalues of the graph Laplacian matrix incrementally. • Capable to estimate CTD in constant time incrementally using the property of random walk and hitting time.	• iECT is faster than iLED but it can only be used in case where a new test point is added and cannot be used when there are weigh updates in the graph. • They take only a small subset of datasets for testing.	-	-
ADWICE TRD [20]	• Performs better on KDDcup99, DARPA98 and safeguard framework datasets.	• ADWICE-TRD explores the BIRCH cluster features for tree formation but it is not feasible in network anomaly detection.	95	2.8
Hybrid IDS [21]	• Capable to detect unknown attacks i.e., the instance of attacks are not used in training.	• They use only DARPA98 dataset. Hence, it needs to be tested over real-life network intrusion datasets.	92.63	1.8
ADWICE Grid [22]	• Performs better than ADWICE-TRD and ADWICE-TRR in terms of detection rate and false alarm rate on real world datasets in safeguard tested.	• They claim that their approach adaptively updates the normal profiles but it is important that the profiles need to be updated based on the different network scenarios.	97.2	1.8

Table 6 (*continued*)

Method	Summary		Performance	
	Pros	Cons	Detection Rate (%)	False Positive Rate (%)
M-ART [23]	• Can work well where both numeric and categorical data are available. • Performs well in terms of detection rate on synthetic and real world UCI datasets.	• Not tested on high dimensional large datasets. • Does not work well where it contains only in numeric data values.	95	2.8
Clonal Selection [24]	• Capable to work in mixed type datasets based their modified objective function. • Performs well in terms of detection rate and false positive rate.	• May not work well in real-time scenarios.	98.3	1.8
System	- simple and quite efficient to detect attacks - can categorized effectively the different types of attacks	- time consuming to find the efficient heuristics rules - among different rules, enable to find most effective heuristics rules	98.5	1.2

7 Evaluation Analysis

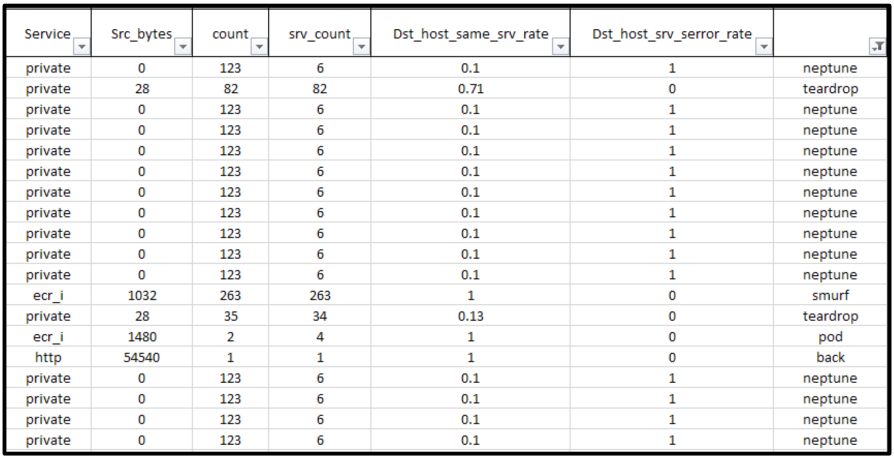
Based on our survey of published papers on incremental anomaly detectors, we observe that most techniques have been validated using the NSL KDD intrusion datasets in an offline mode. However, the effectiveness of an automatic network intrusion detection system based on incremental approach can only be judged in a real-time environment. Following are some of the research issues we have identified in this area.

- Most existing IDSs have been found inadequate with new networking paradigms currently used for both wired and wireless communication 4, 5. Thus, adaptation to new network paradigms needs to be explored.
- Most existing methods are dependent on multiple input parameters. Improper estimation of these parameters leads to high false alarm rates.
- The clustering techniques that are used by anomaly detectors need to be faster and scalable when used on high dimensional and voluminous mixed type data.
- Lack of labeled datasets for training or validation is a crucial issue that needs to be addressed. The KDD99 datasets are out-of-date. New valid datasets need to be created and made available to researchers and practitioners. Developing a reasonably exhaustive dataset for training or validation for use in supervised or semi-supervised anomaly detection approaches is a challenging task.
- Estimation of unbiased anomaly scores for periodic, random or burst attack scenarios is another challenging issue.

- Lack of standard labeling strategies is a major bottleneck in the accurate recognition of normal as well as attack clusters.
- Development of pre or post-processing mechanisms for false alarm minimization is necessary.
- Handling of changing traffic pattern remains a formidable problem to address.

7.1 Denial of Service Attacks (DoS)

This class of attacks makes some computing or memory resource too busy or too full to handle legitimate requests. Thus legitimate users' access are denied. There are several ways to launch DoS attacks such as: by abusing the computers' legitimate features, by targeting the implementations bugs, by exploiting the system's misconfigurations. It is classified based on the services that an attacker renders unavailable to legitimate users. The following table is the evaluation analysis of the DoS attacks using our system.

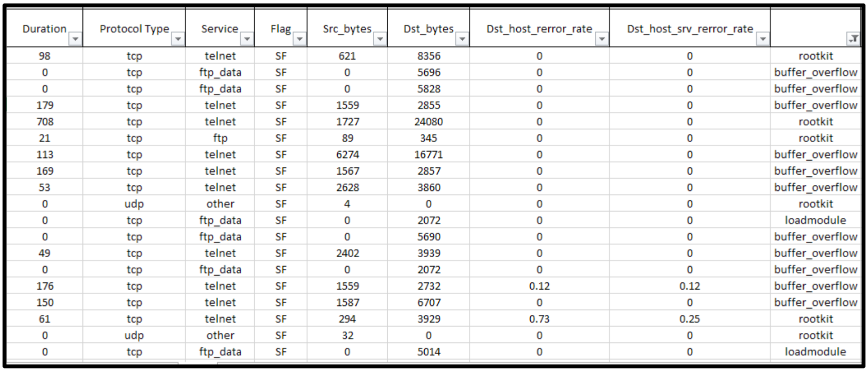


Service	Src_bytes	count	srv_count	Dst_host_same_srv_rate	Dst_host_srv_serror_rate	
private	0	123	6	0.1	1	neptune
private	28	82	82	0.71	0	teardrop
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
ecr_i	1032	263	263	1	0	smurf
private	28	35	34	0.13	0	teardrop
ecr_i	1480	2	4	1	0	pod
http	54540	1	1	1	0	back
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune
private	0	123	6	0.1	1	neptune

Fig. 1 Evaluation Analysis on DoS attacks

7.2 Probing Attacks

Probing is a class of attacks, where an attacker scans a network to gather information or find known vulnerabilities or finds the weaknesses of the machines. An attacker tries to exploit the attacks, when weaknesses of one or more machines have been found. There are different types of probe attacks. Some of them abuse a computer's legitimate features, some of them use social engineering techniques, and some of them use weaknesses of the machine. The following table is the evaluation analysis of the Probing attacks using our system.

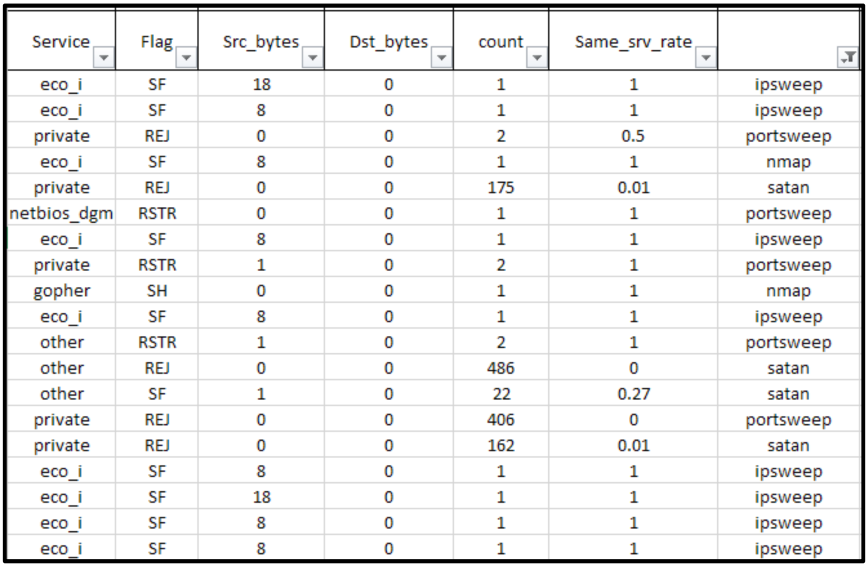


Duration	Protocol Type	Service	Flag	Src_bytes	Dst_bytes	Dst_host_rerror_rate	Dst_host_srv_rerror_rate	
98	tcp	telnet	SF	621	8356	0	0	rootkit
0	tcp	ftp_data	SF	0	5696	0	0	buffer_overflow
0	tcp	ftp_data	SF	0	5828	0	0	buffer_overflow
179	tcp	telnet	SF	1559	2855	0	0	buffer_overflow
708	tcp	telnet	SF	1727	24080	0	0	rootkit
21	tcp	ftp	SF	89	345	0	0	rootkit
113	tcp	telnet	SF	6274	16771	0	0	buffer_overflow
169	tcp	telnet	SF	1567	2857	0	0	buffer_overflow
53	tcp	telnet	SF	2628	3860	0	0	buffer_overflow
0	udp	other	SF	4	0	0	0	rootkit
0	tcp	ftp_data	SF	0	2072	0	0	loadmodule
0	tcp	ftp_data	SF	0	5690	0	0	buffer_overflow
49	tcp	telnet	SF	2402	3939	0	0	buffer_overflow
0	tcp	ftp_data	SF	0	2072	0	0	buffer_overflow
176	tcp	telnet	SF	1559	2732	0.12	0.12	buffer_overflow
150	tcp	telnet	SF	1587	6707	0	0	buffer_overflow
61	tcp	telnet	SF	294	3929	0.73	0.25	rootkit
0	udp	other	SF	32	0	0	0	rootkit
0	tcp	ftp_data	SF	0	5014	0	0	loadmodule

Fig. 2 Evaluation Analysis on Probing Attacks

7.3 User to Roots Attacks (U2R)

In user to root exploits, the attacker starts out with access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system. Most common exploits in this class of attacks are regular buffer overflows, which are caused by regular programming mistakes and environment assumptions. The following table is the evaluation analysis of the U2R attacks using our system.

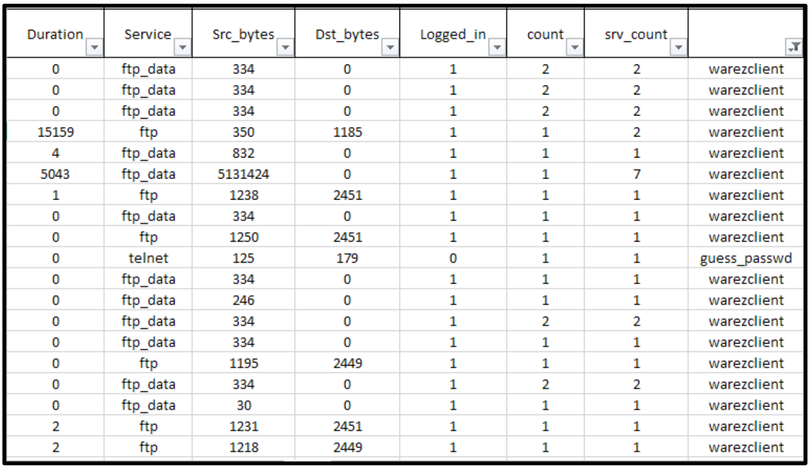


Service	Flag	Src_bytes	Dst_bytes	count	Same_srv_rate	
eco_i	SF	18	0	1	1	ipsweep
eco_i	SF	8	0	1	1	ipsweep
private	REJ	0	0	2	0.5	portsweep
eco_i	SF	8	0	1	1	nmap
private	REJ	0	0	175	0.01	satan
netbios_dgm	RSTR	0	0	1	1	portsweep
eco_i	SF	8	0	1	1	ipsweep
private	RSTR	1	0	2	1	portsweep
gopher	SH	0	0	1	1	nmap
eco_i	SF	8	0	1	1	ipsweep
other	RSTR	1	0	2	1	portsweep
other	REJ	0	0	486	0	satan
other	SF	1	0	22	0.27	satan
private	REJ	0	0	406	0	portsweep
private	REJ	0	0	162	0.01	satan
eco_i	SF	8	0	1	1	ipsweep
eco_i	SF	18	0	1	1	ipsweep
eco_i	SF	8	0	1	1	ipsweep
eco_i	SF	8	0	1	1	ipsweep

Fig. 3 Evaluation Analysis on U2R Attacks

7.4 Remote to Local Attacks (R2L)

A remote to local (R2L) attack makes attack by sending packets to a machine over a network, and then exploits the machines vulnerability to illegally gain local access as a user. There are different types of R2U attacks. The most common attack in this class uses social engineering. The following table is the evaluation analysis of the R2L attacks using our system.



Duration	Service	Src_bytes	Dst_bytes	Logged_in	count	srv_count	
0	ftp_data	334	0	1	2	2	warezclient
0	ftp_data	334	0	1	2	2	warezclient
0	ftp_data	334	0	1	2	2	warezclient
15159	ftp	350	1185	1	1	2	warezclient
4	ftp_data	832	0	1	1	1	warezclient
5043	ftp_data	5131424	0	1	1	7	warezclient
1	ftp	1238	2451	1	1	1	warezclient
0	ftp_data	334	0	1	1	1	warezclient
0	ftp	1250	2451	1	1	1	warezclient
0	telnet	125	179	0	1	1	guess_passwd
0	ftp_data	334	0	1	1	1	warezclient
0	ftp_data	246	0	1	1	1	warezclient
0	ftp_data	334	0	1	2	2	warezclient
0	ftp_data	334	0	1	1	1	warezclient
0	ftp	1195	2449	1	1	1	warezclient
0	ftp_data	334	0	1	2	2	warezclient
0	ftp_data	30	0	1	1	1	warezclient
2	ftp	1231	2451	1	1	1	warezclient
2	ftp	1218	2449	1	1	1	warezclient

Fig. 4 Evaluation Analysis on R2L Attacks

8 System Design

The following is the system design diagram of the research.

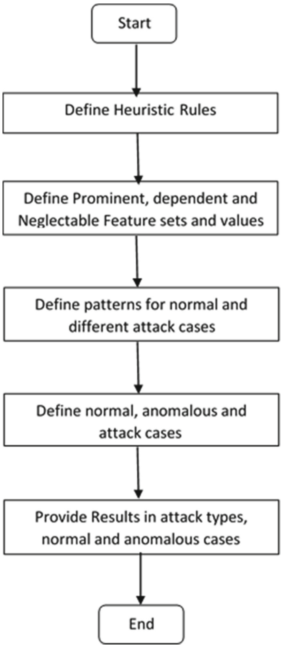

Start
Define Heuristic Rules
Define Prominent, dependent and Neglectable Feature sets and values
Define patterns for normal and different attack cases
Define normal, anomalous and attack cases
Provide Results in attack types, normal and anomalous cases
End

Fig. 5 System Design Diagram

9 Conclusion

As discussed earlier, using the heuristic ruled based anomalous detection technique can provide more efficient and quicker way to understand the nature of different types of attacks. Although defining heuristic rules takes time because it is very difficult to get such confirmed rules without the use of standardized dataset. In this research, we emphasis on standardized network attack detection dataset which is called NSL KDD dataset. So we can entrust the rules that were formulated using this standard dataset.

Earlier research only target to mine and classify the different attack types which is included in NSL KDD dataset. But there is no theoretically testing for the nature of feature set and values changes that can affect the attack or normal pattern. This is the objective of the research to find out the nature and attack pattern. By comparing to other results of previous research, the analysis results seems to be compromising than other previous research because it includes finding out about the nature and pattern of each attack type in detail.

This paper utilized a technique for creating heuristic rules by combining both signature analysis and automated methods. The technique assisted in improving readability, comprehensibility, and maintainability of heuristic rules. Heuristic rules were proposed for five NSL KDD dataset attacks – spy, phf, imap, warezmaster and warezclient. Probability of detection and false alarm rates are computed on the KDD training and testing datasets combined for each considered attack. The overall performance, combined for the three attacks is reasonably good: an average 53.08% detection rate is achieved with only 0.005% false alarm rate for the two specific attacks in the R2L category.

Since the rule thresholds were set using knowledge of both training and testing data sets, the performance of proposed rules on new records in KDD testing data set cannot be interpreted in a precise fashion. On the other hand, signature analysis performed on two R2L attacks is likely to facilitate a predictably good level of detection and false alarm rates for other attacks in the NSL KDD datasets and other type of data derived on real host and network traffic.

References

1. Agarwal, R., Joshi, M.V.: PNrule: A New Framework for Learning Classifier Models in Data Mining (A Case-Study in Network Intrusion Detection). Technical Report TR 00-015, Department of Computer Science, University of Minnesota (2000)
2. Levin, I.: KDD-99 Classifier Learning Contest LLSoft's Results Overview. ACM SIGKDD SIGKDD Explorations **1**(2), 67–75 (2000)
3. Lee, W., Stolfo, S.J., Mok, K.W.: A data mining framework for building intrusion detection models. In: IEEE Symposium on Security and Privacy, Oakland, California, pp. 120–132 (1999)

4. Lindqvist, U., Porras, P.: Detecting computer and network misuse through the production-based expert system toolset (P-{BEST}). In: IEEE Symposium on Security and Privacy, pp. 146–161 (1999)
5. Porras, P.A., Neumann, P.G.: EMERALD: event monitoring enabling responses to anomalous live disturbances. In: Proceedings of the 20th National Information Systems Security Conference, Baltimore, Maryland, pp. 353–365 (1997)
6. Cohen, W.W.: Fast effective rule induction. In: Proceedings of the 12th International Conference on Machine Learning (ML-95), Lake Tahoe, CA: Morgan Kaufmann, pp. 115–123 (1995)
7. DARPA dataset 1998, April 2003. http://www.ll.mit.edu/IST/ideval/data/1998/1998_data_index.html
8. Lee, W., Stolfo, S.J., Mok, K.W.: Mining in a data-flow environment: experience in network intrusion detection. In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, pp. 114–124 (1999)
9. KDD data set, 1999, April 2003. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
10. Yeung, D.Y., Chow, C.: Parzen-window network intrusion detectors. In: Proceedings of the Sixteenth International Conference on Pattern Recognition, Quebec City, Canada, Vol. 4, pp. 385–388, August 2002
11. Lee, W., Stolfo, S.: A Framework for Constructing Features and Models for Intrusion Detection Systems. ACM Transactions on Information and System Security **3**(4), 227–261 (2000)
12. Lippmann, R.P., Haines, J.W., Fried, D.J., Korba, J., Das, K.: The 1999 DARPA Off-Line Intrusion Detection Evaluation. Computer Networks **34**(4), 579–595 (2000)
13. Elkan, C.: Results of the KDD 1999 Classifier Learning. ACM SIGKDD SIGKDD Explorations **1**(2), 63–64 (2000)
14. Yu, W.-Y., Lee, H.-M.: An incremental-learning method for supervised anomaly detection by cascading service classifier and ITI decision tree methods. In: Chen, H., Yang, C.C., Chau, M., Li, S.-H. (eds.) PAISI 2009. LNCS, vol. 5477, pp. 155–160. Springer, Heidelberg (2009). http://dx.doi.org/10.1007/978-3-642-01393-5
15. Laskov, P., Gehl, C., Kr¨uger, S., M¨uller, K.-R.: Incremental support vector learning: Analysis, implementation and applications. Journal of Machine Learning Research **7**, 1909–1936 (2006)
16. Ren, F., Hu, L., Liang, H., Liu, X., Ren, W.: Using density-based incremental clustering for anomaly detection. In: Proceedings of the 2008 International Conference on Computer Science and Software Engineering. Washington, DC, USA, pp. 986–989. IEEE Computer Society (2008). http://dx.doi.org/10.1109/CSSE.2008.811
17. Khreich, W., Granger, E., Miri, A., Sabourin, R.: Adaptive ensembles of HMMs applied to anomaly detection. Pattern Recognition (Elsevier Science), July 19, 2011. doi:10.1016/j.patcog.2011.06.014
18. Yi, Y., Wu, J., Xu, W.: Incremental SVM based on reserved set for network intrusion detection. Journal of Expert Systems with Applications **38**(6), 7698–7707 (2011). USA
19. Lu, N., Khoa, D., Chawla, S.: Online Anomaly Detection Systems Using Incremental Commute Time. CoRR, Vol. abs/1107.3894 (2011)
20. Burbeck, K., Nadjm-Tehrani, S.: ADWICE – anomaly detection with real-time incremental clustering. In: Park, C.-s., Chee, S. (eds.) ICISC 2004. LNCS, vol. 3506, pp. 407–424. Springer, Heidelberg (2005)

21. Rasoulifard, A., Bafghi, A.G., Kahani, M.: Incremental hybrid intrusion detection using ensemble of weak classifiers. In: Sarbazi-Azad, H., Parhami, B., Miremadi, S.-G., Hessabi, S. (eds.) CSICC 2008. CCIS, vol. 6, pp. 577–584. Springer, Heidelberg (2005). doi:10.1007/978-3-540-89985-3
22. Burbeck, K., Nadjm-Tehrani, S.: Adaptive real-time anomaly detection with incremental clustering. Inf. Secur. Tech. Rep. **12**(1), 56–67 (2007). http://dx.doi.org/10.1016/j.istr.2007.02.004
23. Hsu, C.C., Huang, Y.-P.: Incremental clustering of mixed data based on distance hierarchy. Expert Syst. Appl. **35**(3), 1177–1185 (2008). http://dx.doi.org/10.1016/j.eswa.2007.08.049
24. Zhong, C., Li, N.: Incremental clustering algorithm for intrusion detection using clonal selection. In: Proceedings of the 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application. Washington, DC, USA, pp. 326–331. IEEE Computer Society (2008). http://dx.doi.org/10.1109/PACIIA.2008.25

Private Cloud Deployment Model for Academic Environment Using CloudStack

Thinn Thu Nanig

Abstract This paper focuses on the design consideration for private cloud deployment model in academic environment for authors' institution. Conceptual benefits of deploying private cloud system is to provide virtualized computing resources to users rather than accessing physical computing resources. The users of the academic cloud will consist of faculties and students in the institution and they would access computing resource as they need to utilize elastically and flexibly on demand. For deployment model for academic cloud environment, we choose an open source software platform "CloudStack" that pools computing resources to build private cloud as well as to manage the network, storage, and compute nodes for cloud Infrastructure. This system is expected to increase the efficient usage of computing resources, utilization of servers and decreasing the power consumption for physical machines by accessing virtual machines from cloud servers. Another goal is to minimize the operation and maintenance cost for computing resources.

Keywords Academic cloud · CloudStack · Virtualization

1 Introduction

University of Computer Studies, Yangon (UCSY) is the one of the leading universities for ICT academic education in Myanmar. There are more than 4000 students including undergraduate, post graduate and doctorial students are studying in UCSY in each academic year. We designed the computing environment as a centralized system since 2009 by connecting from user's desktop to the servers at anywhere in the campus. More than 500 PCs connected with campus network system including Wi-Fi connection. Every year, operation cost

---

T.T. Nanig(✉)
University of Computer Studies, Taunggyi, Myanmar
e-mail: thinnthu@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_16

and maintenance cost of campus computer are costly. Moreover, a new campus called University of Computer Studies, Yangon (Bahan Campus) has been established in 2012 and more than 350 students are studying in current and about 1000 students will be planned to study in 2017. Now this campus has been changed the name and organization structure as "University of Information Technology" in short as "UIT". The different of courses in UIT is offering degree specialized in high performance computing and Networking technology. Moreover, some related researches in HPC and Networking technology are planned to do in this campus.

For that reason, UIT plans to deploy an academic cloud system to support more computing resources for parallel computing experiments as well as multiple network configuration and installation. By deploying an academic cloud system, the necessary computing resources could be supported to users on demand as virtually and elastically. It leads to be cost effective, easy to maintain the computing servers and networking resources in the campus. Deploying Private Cloud System will support computing resources for university becomes challenging in terms of academic and research. In order to deploy private cloud environment, CloudStack™ is an open source software platform will be used. CloudStack pools computing resources to build public, private, and hybrid Infrastructure as a Service (IaaS) clouds. CloudStack manages the network, storage, and compute nodes that make up a cloud infrastructure. Use CloudStack to deploy, manage, and configure cloud computing environments. Moreover, it has powerful features such as advanced VLAN networking, high availability, and additional network elements such as load balancers and firewalls, and support for multiple hypervisors including Citrix XenServer, KVM, and VMware vSphere. UCSY-Bahan Private Cloud System will be planned to support the virtualized computing resources for HPC Lab and Multimedia Lab.

1.1 Academic Cloud System with VMWare Technology

As the first category, for instance, VMWare technology based academic cloud environment, JAIST private cloud [3,4] is one of the significant examples. Japan Advanced Institute of Science and Technology (JAIST) has deployed an academic cloud system by combining of virtualization products such as VMware vSphere, Citrix XenApp, and Microsoft Application Virtualization with more than 130 servers of Fujitsu PRIMERGY RX200 S2 and BX620 S2. The significant advantage of JAIST private cloud system is that could be saved 48% of power supply and 70% of room space. For that result, JAIST was awarded the green technology by Ministry of the Environments in Japan 2011. However, the total amount of license fees becomes higher. Northwestern University Deploys Private Cloud Infrastructure with VMWare technology is another significant example[5]. Tintri VMstore™ T540 dual-controller 13.5TB storage appliance have been used for each private cloud pod. As the same result like JAIST, license fees becomes increasing according to usage of expensive virtualization products.

1.2 Academic Cloud System with Open Source Technology

Open source cloud technology such as CloudStack[8], OpenStack[9], Eucalyptus[10] have been popular to deploy in academic cloud environment for taking advantage of reducing in operational licensing cost as well as open technology challenges. Stanford university has deployed the Software Defined Networking(SDN) based Private Cloud [7] with OpenFlow technology since 2009. It seems to be very early and pioneer of academic cloud system. More than 6000 servers have been configured for academic cloud system and over 200 campuses could be deployed to access into SDN OpenFlow system. The University of Alabama at Birmingham(UAB) [6] gives scientists a massive, on-demand, virtual storage cloud using OpenStack community by enhancing the Dell cloud solutions. The researchers, faculties in UAB are focused on mass data intensive researches like DNA mapping on high performance cluster to gain new scientific insights by analyzing petabytes (PB) of information. UAB's academic cloud can be achieved more by spending less time managing data.

Hokkaido University has deployed the largest Academic cloud system in Japan using CloudStack on top of the Citrix XenServer hypervisor and is enabling an integrated cloud solution for provisioning, management and operation of IT resources from a unified portal. In addition to creating a more agile, on-demand environment for consuming IT resources, this new cloud offering reduces operational cost and power consumption per server by up to 20 percent, making for a greener IT department[8].

2 Campus Network Design and Private Cloud Design

In this section, there are two detail design of the existing campus network design and proposed academic cloud design will be described in detail. Academic cloud system will be deployed on underlying campus network infrastructure.

2.1 Existing Campus Network Design

Since the campus has already existed the centralized network systems that provides services such as local email service, Moodle learning management system (LMS) service to all users in the campus. When we consider deploying a private cloud system in campus, it will be one of the separate components of the existing network infrastructure. We plan to deploy private cloud system separately apart of the existing network. Then we will connect main switch of private cloud system into campus backbone switch as shown in Figure 1. We plan to configure virtual servers for HPC Lab and Multimedia Lab. In HPC Lab, the students can execute and test run their parallel programs written in Java, OpenMp and MPI on the virtualized parallel computers that instantiate by private cloud system.

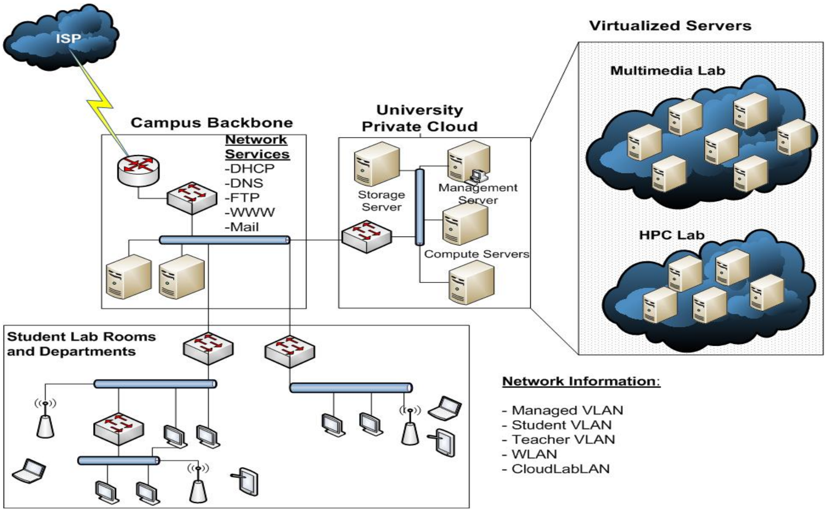

ISP
Campus Backbone
Network Services
-DHCP
-DNS
-FTP
-WWW
-Mail
University Private Cloud
Storage Server
Management Server
Compute Servers
Virtualized Servers
Multimedia Lab
HPC Lab
Student Lab Rooms and Departments
Network Information:
- Managed VLAN
- Student VLAN
- Teacher VLAN
- WLAN
- CloudLabLAN

Fig. 1 General Network Design of UIT and Private Cloud System.

2.2 *Logical Design for Academic Cloud System*

Private Cloud Computing Lab will be deployed underlying University Network backbone. Cloud Computing Lab will instantiate virtual machines which are running on underlying deployed private cloud system to be accessed by students in UIT. The logical design of the private cloud system is showed in figure 2.

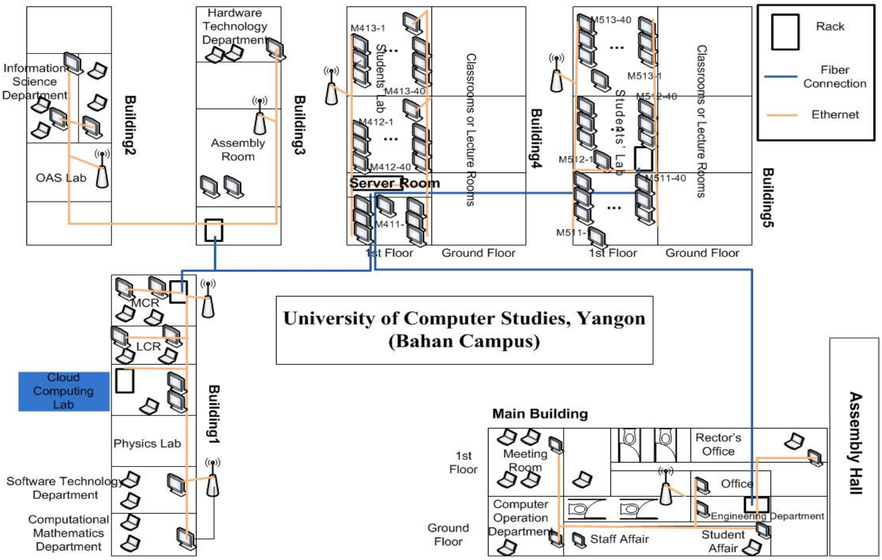

Information Science Department
OAS Lab
Building2
Hardware Technology Department
Assembly Room
Building3
M413-1
M413-40
M412-1
M412-40
M411-
Students' Lab
Server Room
Classrooms or Lecture Rooms
1st Floor
Ground Floor
Building4
M513-40
M513-1
M512-40
M512-1
M511-40
M511-
Students' Lab
Classrooms or Lecture Rooms
1st Floor
Ground Floor
Building5
Rack
Fiber Connection
Ethernet
MCR
LCR
Cloud Computing Lab
Physics Lab
Software Technology Department
Computational Mathematics Department
Building1
University of Computer Studies, Yangon (Bahan Campus)
Main Building
1st Floor
Ground Floor
Meeting Room
Rector's Office
Office
Engineering Department
Computer Operation Department
Staff Affair
Student Affair
Assembly Hall

Fig. 2 Logical design of Private Cloud System

The users of the academic cloud will consist of faculties and students in the campus and they would access computing resource as they need to utilize elastically and flexibly on demand. Especially, both faculties and students who need to utilize more computing resources to test parallel computing programs for mathematical applications, multimedia applications. They could access virtual machines through private cloud system as they need.

3 Private Cloud System Deployment

According to the study of successful academic cloud deployment models as mentioned in Section 3, we choose open source “Apache CloudStack” framework to deploy our institute’s academic could system. There are two strong reasons why we choose open source systems: first, we would minimize the cost for installation, operation and maintenance of service providing by academic institution; second reason is to initiate the open source technology as well as for combining and integrating of open cloud frameworks such as Hadoop and ClouStack. Apache CloudStack is open source platform that offers a complete platform to set up a Private or Hybrid cloud for enterprises as well as academic institution.

3.1 Apache CloudStack Architecture

Then we will introduce the fundamental components of CloudStack and detail description and their responsibility are shown in Table1. There are several components have to be deployed into CloudStack private cloud platform: Management servers, CloudDB, Compute Node, Primary Storage, Secondary Storage. As description in Table 1, the main responsibility of management server is to manage the overall cloud infrastructure: Cloud administration, guest VM assignment, private and public IP assignment, and to allocate virtual disk storage. CloudDB is primary DB server to be set up a MySQL cluster to be ensured high availability. The basic units of CloudStack management server are: (i) The User Interface and Application Program interface.(ii) the business logic (iii) Orchestration Engine and (iv) controller.

Table 1 Description of CloudStack components

Component	Descriptions and responsibility
Management Server	It manages overall infrastructure and resources in CloudStack. It provides (i) the web user interface for both the administrator and the users, (ii) assignment of guest VMs to particular hosts, (iii) IP address assignment for both public and private and (iv) allocate the virtual disk storage.

Table 2 *(continued)*

Component	Descriptions and responsibility
CloudDB	The CloudDB is the primary MySQL database. CloudDB is setting up a MySQL cluster to be ensured High availability.
Compute Node	Compute node is the host where Hypervisors have been resided.
Primary Storage	Primary storage is used to host the guest virtual machines in the cluster stores. The volume of these virtual machines and the allocation of guest virtual disks to the primary storage.
Secondary Storage	This storage space is used to store the templates, ISO images and snapshots that can be used to deploy IT infrastructure using CloudStack.
Zone	Zones are the highest level of hierarchy in which the datacenter of an organization is logically partitioned to provide physical isolation and redundancy.
Pod	A pod is the second level of hierarchy in the deployment of the CloudStack. Pod defines the management subnet of the system VMs by the management server.
Cluster	A cluster is the third level of hierarchical division in the deployment of CloudStack. It can also be "live migrated" to another host in the same cluster using a shared storage device.

3.2 Design Consideration for Cloud Deployment Model in the Campus

In order to deploy an academic cloud infrastructure in the campus, there are three physical servers have been deployed as Management and MySQL server, Computer Node and Primary storage server and Secondary storage servers specification are described in Table 2 and Table 3. We have used three server components to be configured for CloudStack Infrastructure, they are : (i) one PC server for Management Server and CloudDB configuration; (ii) powerful dell PowerEdge R920 server for Computer Node server (in which all virtual machines will be runned), primary storage server (in which all VM images and storage volumes are stored) are configured; (iii) Synology NAS storage will be configure as Secondary Storage server. The detail deployment design of Academic Cloud System is showed in figure 3.

Table 2 Specification for Server (Computer Node and Primary Storage)

Brand	PowerEdge R920
Professor	2 CPU x Intel Xeon E7-4809 v2 Processor 1.9GHz, 12M Cache, 6.4 GT/s QPI, No Turbo, 6 Core, 105W
Memory	128GB (8 X 16GB RDIMM, 1600 MHz, Low Volt, Dual Rank, x4)
HDD	10Nos.x 1TB 7.2K RPM, 6Gbps Near Line SAS 2.5" Hard Drive - HotPlug
Controller	PERC H730P Adapter RAID Controller, 2GB NV Cache

Table 3 Specification for Server (Secondary Storage)

Brand	Synology NAS (8 Bays)
Professor	Dual Core 2.13GHz CPU
Memory	2 x 2GB DDR3 RAM
HDD	4 x Western Digital 1TB (Red) Hard Disk

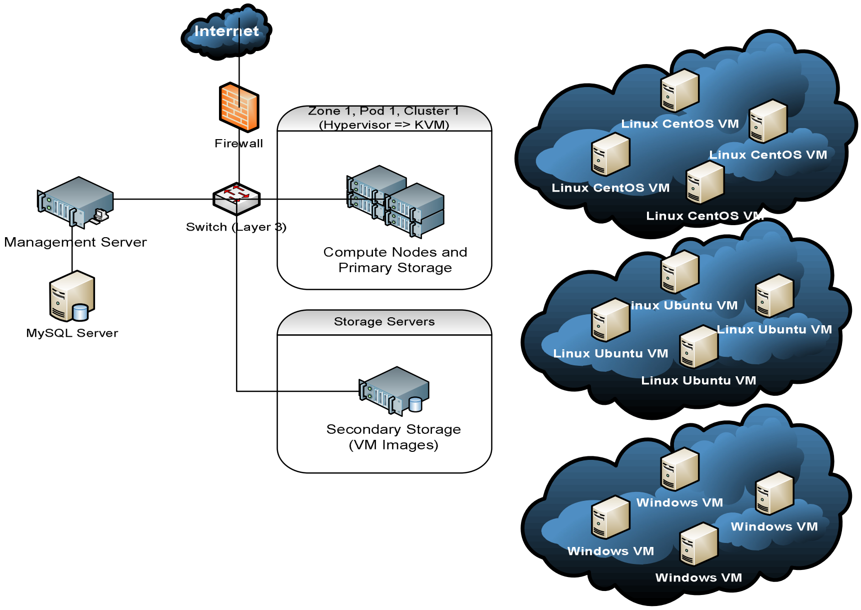

Internet
Firewall
Switch (Layer 3)
Management Server
MySQL Server
Zone 1, Pod 1, Cluster 1
(Hypervisor => KVM)
Compute Nodes and
Primary Storage
Storage Servers
Secondary Storage
(VM Images)
Linux CentOS VM
Linux CentOS VM
Linux CentOS VM
Linux CentOS VM
Linux Ubuntu VM
Linux Ubuntu VM
Linux Ubuntu VM
Linux Ubuntu VM
Windows VM
Windows VM
Windows VM
Windows VM

Fig. 3 Physical design model of CloudStack Infrastructure

3.3 Software Design and Deployment Plan

As previously mentioned in above section, UIT Academic Cloud System will be deployed using Open Source Platform Apache CloudStack. Table 4 shows the software specification for CloudStack deployment:

Table 4 Software Components

Server Component	Operating System and Software	Remarks
Management Server CloudDB Server Compute Node	Operating System : CentOS 6.5 x86 Other Server : MySQL Server Cloud Server : CloudStack version 4.4.1 Virtualization : KVM hypervisor Server	(1) Configure the network (2) Host name (3) SELinux (4) NTP (5) Configuring the CloudStack Package Repository
Primary Storage Server	Operating System: CentOS 6.5 x86 File Server : NFS Server (NFSv4)	
Secondary Storage Server	Operating System: CentOS 6.5 x86 File Server : NFS Server (NFSv4)	

In this research, we have chosen Apache CloudStack release 4.4.1 on underlying CentOS Linux operating system version 6.5 as well as we choose KVM hypervisor for virtualization technology. For both primary and secondary storage servers installation, we choose NFS file system version 4 on CentOS version 6.5.

4 Implementation Plans

Phase 1 (Procurement and Installation): To deploy academic cloud system using open source systems (CloudStack) to provide IaaS (Infrastructure as a Service) which can instantiate virtualized computing resources including virtual machines, virtual switches, and virtual storage and so on. In this phase, equipment procurement will be prepared and private cloud installation and configuration will be done.

Phase 2 (Testing): To establish virtualized labs which can be involved to provide computing services for particular subjects such as application implementation and testing for Software Engineering, High Performance Computing, Mobile Computing, Distributed System and Multimedia System. In this phase, the necessary surveys and researches will be done.

Phase 3 (Implementing and Evaluation): This phase will evaluate the efficiency and effectiveness of academic cloud system and virtualized labs.

Phase 4 (Research Presentation): This phase will present the result and outcome of academic cloud system and virtualized labs.

5 Usability of the Cloud System

According to configured servers and hardware specification, we plan to instantiate more than 50 virtual machines at the same time, it means that maximum 50 students could access the cloud system simultaneously. It is easy to extend by attached with extra compute node server as well as clusters and zones. We expect to enhance for more cluster nodes configuration into the existing cloud system after testing processes have been completed. Moreover, system testing and performance testing will be planned after component testing as well.

6 Advantage and Disadvantage of ClockStack Based Private Cloud System

There are several advantages of deploying private cloud system in the campus: (i) first, the students could gained the knowledge and experience to access cloud system even in their university campus; (ii) the related researches of Cloud Computing such as *Cloud Services*, *Cloud Storage* and *Cloud Security* could be tested on the existing private cloud system; (iii) the experience gained by deploying private cloud system can be applied in other applications such as e-Government, e-Banking and other universities; (iv) *the green technology* could be realized by accessing virtual computing resources rather than accessing physical computers; (v) Moreover, we could reduce installation and maintenance cost by using Open source CloudStack Platform rather than utilization of expensive licensed version cloud product like VMWare; (vi) CloudStack is also compatible to integrate with other Cloud venders such as VMWare and XenServer.

On the other hand, there are several disadvantages in deploying CloudStack based academic cloud system; (i) VM images and virtual machines on CloudStack are only relied on Linux OS, for that reason, Virtual Machines provided on Virtual Labs are only available Linux distribution OS and students will be encouraged to utilize Linux VM for their testing environment;(ii) Since we choose open source Cloud platform, several issues will be introduced in terms of Security that is critical challenges to be solved as other related researches; (iii) Unlike expensive Cloud product (e.g. VMWare), lack of services such as Cloud Migration that should be enhanced as another related researches.

7 Conclusion

UIT now has started to deploy an academic cloud system to support more computing resources that could be supported to users on demand as virtually and elastically. This is pilot research to leads for cost effective, easy to maintain the computing servers and networking resources in the campus. As we mentioned in previous section, there are several advantages gained in this research: cloud research extensibility, utilized experience in other projects, realizing green technology, cost and effectiveness in terms of hardware and software, then easy to integrate with other cloud system.

Acknowledgement This research is supported by Asia Pacific Telecommunity(APT) HRD and Exchange Project together with University of Information Technology (UIT) , KDDI Foundation, Japan and Waseda University Japan. The equipments used in this project are funded by APT Project in 2014-2015. The author has worked in UIT during the period of that project. Now the author has moved to University of Computer Studies, Taunggyi in Shan State.

References

1. Kureshi, I., Pulley, C., Brennan, J., Holmes, V., Bonner, S., James, Y.: Advancing Research Infrastructure Using OpenStack. (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Extended Papers from Science and Information Conference, 63–69 (2013)
2. Barker, A., Varghese, B., Ward, J.S. Sommerville, I.: Academic cloud computing research: five pitfalls and five opportunities. In: 6th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 14) (2014)
3. Shikida, M., Nakano, H., Kozaka, S., Mato, M., Uda, S.: A centralized storage system with automated data tiering for private cloud environment. In: Proceedings of the 41st Annual ACM SIGUCCS Conference on User Services, November 2013
4. Shikida, M., Miyashita, K., Ueno, M., Uda, S.: An evaluation of private cloud system for desktop environments. In: SIGUCCS 2012 Proceedings of the 40th Annual ACM SIGUCCS Conference on User Services, pp. 131–134 (2012)
5. Northwestern University Deploys Private Cloud Infrastructure with Tintri. IT White Paper (2012). http://tintri.co.jp/sites/default/files/field/pdf/document/tintri-casestudy-northwestern-140506.pdf
6. Robinson, J.-P.: University uses cloud to unleash research potential. https://www.uab.edu/it/home/images/Case_Study_University_of_Alabama_Uses_Cloud_to_Unleash_Research_Potential.pdf
7. van Reijendam, J.: Stanford SDN-Based Private Cloud. http://www.cenic.org/wp-content/uploads/2013/05/JohanvanReijendam.pdf
8. Apache CloudStack Cloud Computing
9. OpenStack. http://www.openstack.org/
10. Eucalyptus: Open Source Private Cloud Software. https://www.eucalyptus.com/eucalyptus-cloud/iaas
11. Hadoop: The Definitive Guide (Book)

Two Methods for Color Constancy Based on the Color Correlation Matrix

Takashi Toriu, Mikiya Hironaga and Naoya Hasebe

Abstract Two methods for color constancy are proposed in this paper. Both are based on the correlation matrix on the three-dimensional space of colors, red, green and blue. In the first method, the eigenvector corresponding to the largest eigenvalue is assumed to be a good estimate of the illumination color, and the influence of the illumination color is eliminated from the input image. In the second method, it is assumed that the eigenvector corresponding to the largest eigenvalue presents the color gray when the illumination is white The image under white illumination is obtained by a iteration method so as to satisfy the condition that the eigenvector corresponding to the largest eigenvalue presents the color gray. These two methods were compared to a widely used typical method for color constancy, the Gray-world method, by simulation experiments using synthesized images and real images, and the effectiveness of the proposed methods was validated by these experiments.

1 Introduction

Color recognition has been being an important and basic task in computer vison. In fact, color recognition has one serious problem that a small change of the illumination condition strongly affects the recognition accuracy and the color recognition often fails. On the other hand, human being perceives the object color independent of the illumination [1]. This ability is called color constancy. In this paper, we discuss how to realize this ability on computer vision.

There have been proposed several methods for color constancy. The papers [2, 3] proposed methods to predict illumination color based on the relation between specula reflection and diffusion reflection. In the papers [4-6], methods based on

---

T. Toriu(✉) · N. Hasebe
Graduate School of Engineering, Osaka City University, Osaka, Japan
e-mail: toriu@info.eng.osaka-cu.ac.jp

M. Hironaga
School of Science and Engineering, Kinki University, Osaka, Japan

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_17

Gray-world assumption were proposed. In these methods, the influence of the illuminant is eliminated by assuming the average of the colors of the objects in the scene is gray. This method would work well when sufficient colors exist in the scene. A method based on local averaging was proposed in the paper [7]. Further, there have been proposed methods using image statistics such as correlation between the brightness and color [8-12]. The paper [8] discussed human color constancy based on the correlation between luminance and redness and concluded that redness of the illumination could be correctly estimated using the mean redness of the image and the correlation between luminance and redness of the image. Inspired by the paper [8], we developed a computational method to estimate illumination color by replacing human eyes with a camera [11]. In the previous paper [13], we proposed another method based on "Minimum Brightness Variance Assumption", in which it is assumed that variance of brightness is minimum when the illumination is white.

In this paper, we intend to combine both methods of Gray-world assumption based methods and correlation based methods. More concretely, we propose two methods for color constancy. Both are based on the correlation matrix on the three-dimensional space of colors, red, green and blue. In the first method, the eigenvector corresponding to the largest eigenvalue is assumed to be a good estimated of the illumination color, and the influence of the illumination color is eliminate from the input image. In this method, the eigenvector corresponding to the largest eigenvalue plays the role of the average color in the method based on Gray-world assumption. In the second method, it is assumed that the eigenvector corresponding to the largest eigenvalue presents the color gray when the illumination is white. The image under white illumination is predicted by an iteration so as to satisfy the condition that the eigenvector corresponding to the largest eigenvalue presents the color gray. These two methods were compared to the Gray-world assumption based method, by simulation experiments using synthesized images and real images, and effectiveness of the proposed methods was validated by these experiments.

2 Basic Idea

The method based on "Minimum Brightness Variance Assumption" [13] sometimes fails to correctly estimate illumination color, because it is not always true that variance of brightness is minimum when illumination is white. In this paper, we are discussing new methods to overcome this problem.

Let $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ be three unit vectors representing color orientations of red, green and blue in the three-dimensional color space. Then the unit vector $\mathbf{e}^{(L)} = 1/\sqrt{3}(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)$ represents the orientation of the white color. It is represented as $\mathbf{e}^{(L)} = 1/\sqrt{3}\,(1, 1, 1)^t$ in terms of components. You can say that this vector also represents the orientation of luminance. Let $I_1(x, y)$, $I_2(x, y)$

and $I_3(x, y)$ be the color components red, green and blue of the input image at the point (x, y), respectively. This image can be represented as

$$\mathbf{I}(x, y) = I_1(x, y)\mathbf{e}_1 + I_2(x, y)\mathbf{e}_2 + I_3(x, y)\mathbf{e}_3, \tag{1}$$

in the three-dimensional color space. The luminance of the image is expressed as

$$L(x, y) = (\mathbf{I}(x, y), \mathbf{e}^{(L)}) = 1/\sqrt{3}\left(I_1(x, y) + I_2(x, y) + I_3(x, y)\right). \tag{2}$$

Then the mean luminance $< L >$ is defined as

$$< L >= \frac{1}{N} \sum_{x,y} L(x, y), \tag{3}$$

and the covariance of the luminance $C^{(L)}$ is defined as

$$C^{(L)} = \frac{1}{N} \sum_{x,y} \left(L(x, y) - < L >\right)^2, \tag{4}$$

where N is the number of pixels. Let the covariance matrix C be

$$C = \frac{1}{N} \sum_{x,y} (\mathbf{I}(x, y) - < \mathbf{I} >)(\mathbf{I}^t(x, y) - < \mathbf{I}^t >), \tag{5}$$

where $< \mathbf{I} >$ is the mean of $\mathbf{I}(x, y)$. Then we can easily prove that $C^{(L)}$ can be expressed as

$$C^{(L)} = (\mathbf{e}^{(L)}, C\mathbf{e}^{(L)}). \tag{6}$$

In fact, if we substitute (2) into (4), we obtain (6). In the case that “Minimum Brightness Variance Assumption” is true, (6) implies that $\mathbf{e}^{(L)}$ would be the eigenvector of C corresponding to the smallest eigenvalue if illumination is white. In this context, it is valuable to note that the eigenvector corresponding to the smallest eigenvalue is unstable when the covariance matrix C is closes to isotropic, because three eigenvectors have the same eigenvalues when C is isotropic. This would be the reason why the method based on “Minimum Brightness Variance Assumption” sometimes fails.

On the other hand, the correlation matrix is far from isotropic. The correlation matrix K is defined as

$$K = \frac{1}{N} \sum_{x,y} \mathbf{I}(x, y)\mathbf{I}^t(x, y). \tag{7}$$

Because the components of $\mathbf{I}$ are always positive, the correlation matrix K is in fact far from isotropic. Then, it is expected that the eigenvector corresponding to the largest eigenvalue is rather stable against changes in the illumination condition. Considering this fact, we developed new methods for color constancy using the correlation matrix.

3 Algorithms

3.1 The First Method (Method 1)

Let $\mathbf{u}$ be the eigenvector of the correlation matrix K corresponding to the largest eigenvalue λ. Assuming that $\mathbf{u}$ represents the orientation of the illumination color, we estimate the color components $\hat{I}_1(x,y)$, $\hat{I}_2(x,y)$, $\hat{I}_3(x,y)$ of the input image under white illumination as

$$\hat{I}_i(x,y) = \alpha I_i(x,y)/u_i, \tag{8}$$

where $\mathbf{u} = (\mathrm{u}_1, \mathrm{u}_2, \mathrm{u}_2)^{\mathrm{t}}$ and α is

$$\alpha = \frac{I_1(x,y)+I_2(x,y)+I_3(x,y)}{I_1(x,y)/u_1+I_2(x,y)/u_2+I_3(x,y)/u_3}, \tag{9}$$

which is determined to keep the brightness invariant.

3.2 The Second Method (Method 2)

Let $\hat{I}_1(x,y)$, $\hat{I}_2(x,y)$, $\hat{I}_3(x,y)$ be the color components of the image to be estimated and let $\tilde{K}$ be the correlation matrix of these. $\tilde{K}$ is obtained as

$$\tilde{K} = \frac{1}{N}\sum_{x,y}\hat{\mathbf{I}}(x,y)\hat{\mathbf{I}}^t(x,y). \tag{10}$$

We assume that the eigenvector of $\tilde{K}$ corresponding to the largest eigenvalue is parallel to the orientation of color gray when the illumination is white. The image corresponding to the white illumination is estimated to find such $\hat{\mathbf{I}}(x,y)$ that satisfy the above assumption. In other words, $\hat{\mathbf{I}}(x,y)$ is estimated so that $\mathbf{e}_l = 1/\sqrt{3}\begin{pmatrix}1 & 1 & 1\end{pmatrix}^t$ is the eigenvector of $\tilde{K}$ with the largest eigenvalue. More specifically, $\hat{\mathbf{I}}(x,y)$ is obtained by following procedure.

(1) Initialization: We set

$$\hat{I}_i(x,y) = I_i(x,y) \text{ for } i = 1,2,3, \tag{11}$$

where $\mathbf{I}(x,y)$ is the input image.

(2) Updating: Calculate $\tilde{K}$ according to the equation (10), and obtain the eigenvector $\mathbf{u}$ of $\tilde{K}$ with the largest eigenvalue. Update $\hat{\mathbf{I}}(x, y)$ according to

$$\hat{I}_i(x, y) \leftarrow \alpha \frac{\beta u_i + 1}{2u_i} \hat{I}_i(x, y), \tag{12}$$

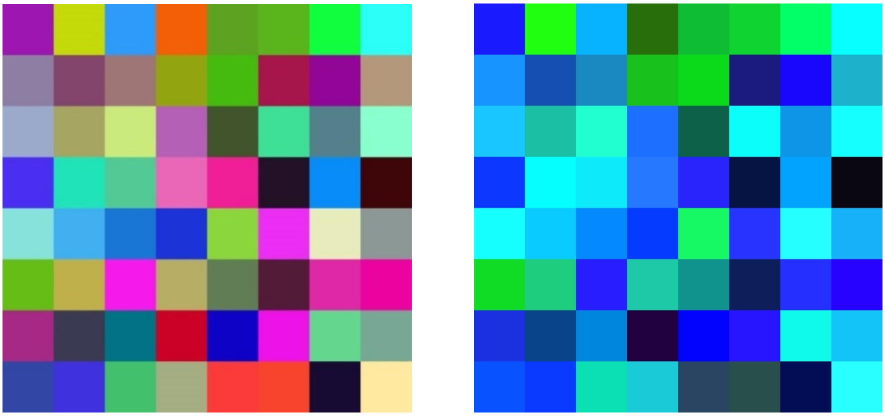
where α is determined to keep the brightness invariant, and β is a factor to modulate the speed of the image change. The larger β is, the smaller the image change is.

(3) Termination condition:

The above procedure should be terminated when the amount of image change according to (12) is less than a pre-determined value. The amount of image change is evaluated by the rout-mean-square error (RMSE).

4 Experiments

We conducted two types of simulation experiments to confirm the effectiveness of the proposed methods. In the first type experiments, first, we synthesized images with 512 × 512 size which represents the object color. The image has 8 × 8 = 64 blocks, and in each block the object colors are specified by $S_1(x, y)$, $S_1(x, y)$ and $S_3(x, y)$, which are determined at random and uniformly distributed from 0.0 to 1.0. In this manner, we prepared 100 samples of images representing object colors. An example is shown in Fig. 1 (a).



(a) An example of images presenting object color.

(b) An example of input images.

Fig. 1 Examples of synthesized images for evaluation.

Next, we set E_1, E_2, E_3 to be the color components red, green, blue of the illumination, and we generated input images as

$$I_i(x, y) = E_i S_i(x, y) \ for \ i = 1,2,3 . \tag{13}$$

Each E_1, E_2, E_3 is set randomly to be one of four values 0.1, 0.4, 0.7 and 1.0. We prepared 20 kinds of illumination color “r, g, b” as is shown in Fig. 2. For each kind of illumination color we prepared 100 images using equation (13). Thus, we prepared 2,000 images in all. An example of generated image is shown in Fig.1 (b).

For these 2,000 images we estimated images under white illumination based on two proposed methods. We also estimated the images using the Gray-world based method for comparison. In Fig. 2, the root-mean-square error between the ideal image and the estimated images of the Gray-world based method, Method 1 and Method2 are shown. Examples of estimated imaged by the two proposed methods are shown in Fig.3. We can see that Method2 has consistently better results than other methods.

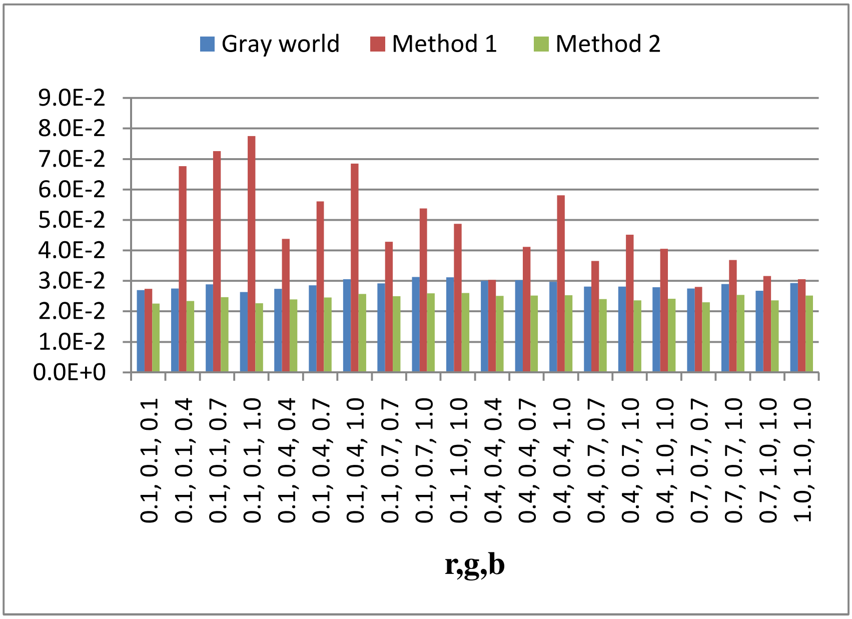
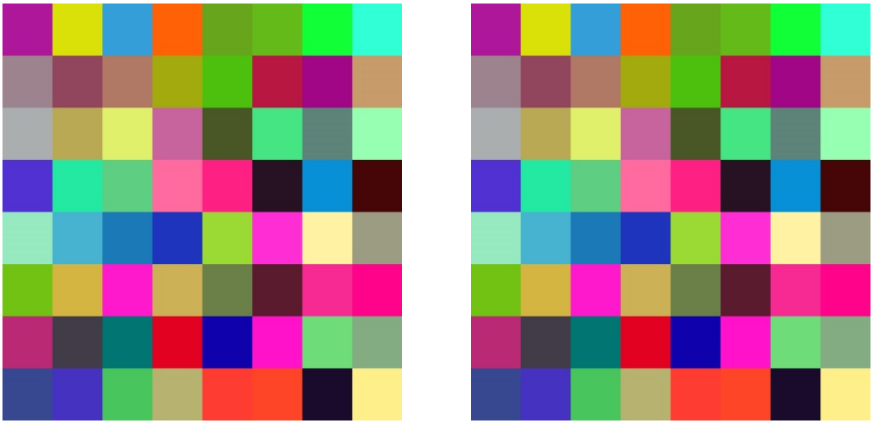


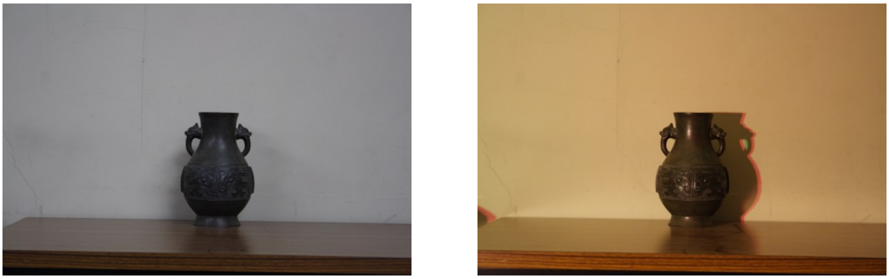
Fig. 2 Root-mean-square error between the ideal image and the estimated image in the experiments using synthesized images.



(a) The estimated image by Method1.

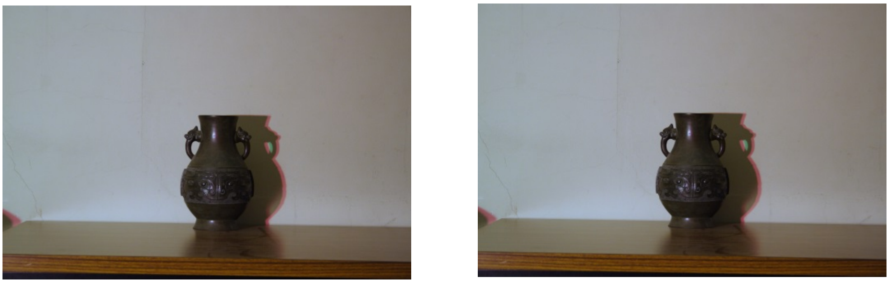
(b) The estimated image by Method 2.

Fig. 3 Examples of estimated images by the two proposed methods.



(a) An input image IM1 under white illumination.

(b) An input image IM1 under colored illumination.



(c) The estimated image by Method1.

(d) The estimated image by Method2.

Fig. 4 Examples of input images and estimated images by the two proposed methods.

Next, we conducted experiments using 18 real images. Fig. 4 (a) and (b) show an example of the input image under white illumination and colored illumination, respectively. (c) and (d) show the estimated images under white illumination by Method 1 and Method 2, respectively. Fig. 4 shows the root-mean-square error between the ideal image and the estimated images of the Gray-world based method, Method 1 and Method 2. For these real images, Method 1 and Method 2 have comparative efficiency and consistently better than the Gray-world based method.

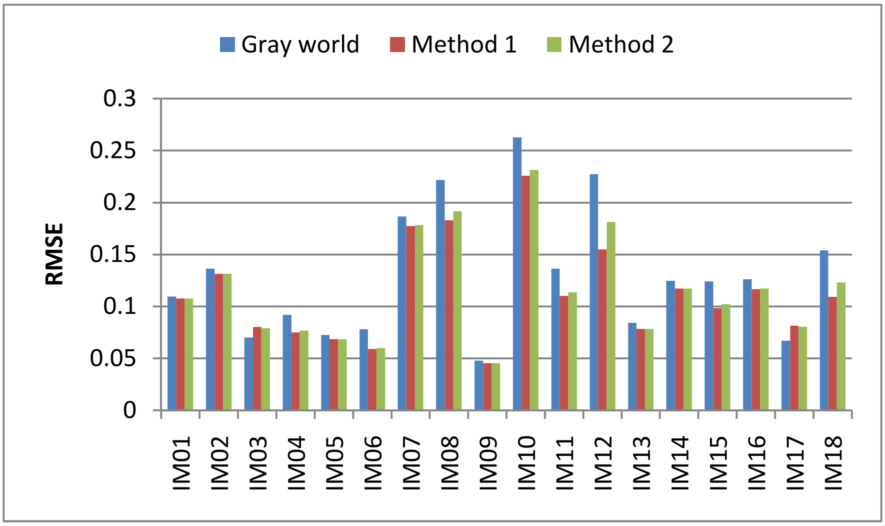


Fig. 5 Root-mean-square error between the ideal image and the estimated image in the experiments using real images.

5 Conclusions

We proposed two methods for color constancy based on the correlation matrix on the three-dimensional space of colors, red, green and blue. In the first method, the eigenvector corresponding to the largest eigenvalue is assumed to be a good estimate of the illumination color. In the second method, it is assumed that the eigenvector corresponding to the largest eigenvalue presents the color gray when the illumination is gray. We showed the second method has consistently advantage compared to the Gray-world based method, by simulation experiments using synthesized images and real images.

References

1. Foster, D.H.: Color constancy. Vision Research **51**, 674–700 (2011)
2. Shafer, S.A.: Using color to separate reflection components. COLOR Research and Application **10**(4), 210–218 (1985)
3. Zmura, M.D., Lennie, P.; Mechanisms of color constancy. J. Opt. Soc. Am. A **3**(10), October 1986
4. Barnard, K., Cardei, V., Funt, B.: A comparison of computational color constancy algorithms. Part I: Methodology and experiments with synthesized data. IEEE Trans. IP **11**(9), 972–984 (2002)
5. Kawamura, H., Fukushima, K., Sonehara, N.: Mathematical conditions for estimating illumination color based on gray world assumption. In: Proc. of the Society Conference of IEICE, vol. 167 (1995)
6. Ciurea, F., Funt, B.: A large image database for color constancy research. In: IS&T/SID Eleventh Color Imaging Conference (2003)
7. Gijsenij, A., Gevers, T.: Color constancy by local averaging. In: 14th International Conference of Image Analysis and Processing, Workshops (2007)
8. Mac.Leod, G.: Influence of scene statistics on color constancy. Nature **415**, 637–640 (2002)
9. Golz, J.: The role of chromatic scene statistics in color constancy: Spatial integration. Journal of Vision **8**(13), 6:1–6:16 (2008)
10. Barnard, K., Martin, L., Funt, B.: Colour by correlation in a three-dimensional colour space. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1842, pp. 375–389. Springer, Heidelberg (2000)
11. Yoshida, T., Hironaga, M., Toriu, T.: Estimating and Eliminating a Biased Illumination with the Correlation between Luminance and Colors. ICIC Express Letters: An International Journal of Research and Surveys **7**, 1687–1692 (2012)
12. Granzier, J.J.M., Brenner, E., Cornelissen, F.W., Smeets, J.B.J.: Luminance color correlation is not used to estimate the color of the illumination. Journal of Vision **5**, 20–27 (2005)
13. Hasebe, N., Hironaga, M., Toriu, T.: A color constancy model with minimum brightness variance assumption. In: ICIVC (2014)

Machine Learning Techniques for Myanmar Word-Sense Disambiguation

Phyu Phyu Khaing and Than Nwe Aung

Abstract Word Sense Disambiguation (WSD) is the vital of Natural Language processing such as machine translation, grammatical analysis, content analysis and information retrieval. WSD process is useful for automatically identifying the correct meaning of an ambiguous word in the sentence or the query when it has multiple meanings. In this paper, the supervised, semi-supervised, unsupervised and knowledge-based approaches for WSD are discussed. This work aim to explore the machine learning techniques for word sense disambiguation of Myanmar Nouns.

Keywords Word sense disambiguation · Natural language processing · Machine learning

1 Introduction

Myanmar language is the member of Tibeto-Burman language and is the sub family of Sino-Tibetan family of language. It is the official language of Myanmar and it also known as Burma. It as Southeast Asia language do not use white space between words. The sentence order is subject+object+verb. Myanmar language is script shaped in circular and semi-circular letters and is adopted from Mon script. There are 33 consonants and 12 vowels in Myanmar language.

Natural language processing (NLP) is an application and area of research that explores the machine translation of human languages to useful things. The text or speech of language need to manipulate. So text based natural language processing has been regarded as consisting of different levels, they are;

P.P. Khaing(✉) · T.N. Aung
University of Computer Studies, Mandalay, Myanmar
e-mail: {Phyuphyukhaing07,mdytna}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_18

Lexical Analysis: analysis of word forms.
Syntactic Analysis: structure processing.
Semantic Analysis: analysis of meaning.
Discourse Analysis: processing interrelated sentences.
Pragmatic Analysis: The purposeful use of sentences in situations.

Ambiguity can occur at all these levels. Ambiguity can be referred to as having more than one meaning/ sense or can be understand more than one way. So computer cannot understand the language like people do.

1.1 Word-Sense Ambiguity

The ambiguity is major part of any human language. Most of the words of human languages uses polysemous that is it have variety of meaning of sentences. There are majorly different types of ambiguity, they are;
Lexical Ambiguity
Syntactic Ambiguity
Semantic Ambiguity
Discourse Ambiguity
Pragmatic Ambiguity

Word sense disambiguation is the task of considering the meaning of an ambiguous word in its context and it is important task in natural language processing. WSD is complex task that is not useful by itself. But it is key task to natural language processing such as;

- Information Retrieval
- Machine Translation
- Location Finding
- Question and answering
- Knowledge acquisition
- Corpus tagging

2 Word-Sense Disambiguation (WSD)

The word-sense disambiguation aims at determining the appropriate meaning of word in a particular context and it is highly related with natural language processing. The resolving the ambiguity of the word is main problem of the understanding of natural language processing. So automatic system which can find out the correct meaning of word in a particular context. There are two main approaches in WSD, deep approach and shallow approach. Deep approach uses the some kind of knowledge that relates to the words and shallow approach sense the context in which the word has been used. The conceptual model of WSD is as follows;[12]

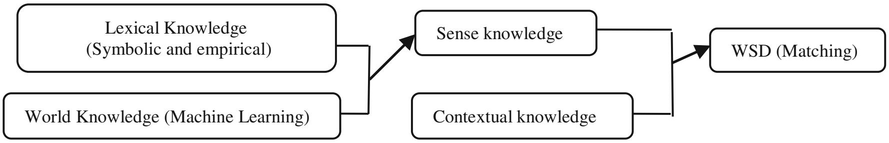

Lexical Knowledge (Symbolic and empirical)
World Knowledge (Machine Learning)
Sense knowledge
Contextual knowledge
WSD (Matching)

Fig. 1 Conceptual model of Word Sense Disambiguation

WSD is as hard as most of the difficult problem in artificial intelligence. So WSD is also regarded as Artificial Intelligence (AI) complete problem. WSD can improve the process including the associations of tasks including machine translation, information retrieval and hypertext navigation, parsing, speech synthesis, spelling correction, reference solution, automatic text summarization, etc.

2.1 *Difficulties of Word-Sense Disambiguation*

The many researcher define the most three difficulties of WSD are as follows.

Discreteness of the Sense. Discreteness of the sense relates with the level of distinction that a sense should have for determining it from a different sense of a word. The concept of word sense is ambiguous causing disagreement among lexicographers. There are two levels of discreteness of the sense they are coarse-grained and fine-grained. In coarse-grained level deals with homograph which is a word have different meanings. It can easily distinguish by human. In the other hand, the fine-grained level is difficult even for human. It occurs for the WSD use different dictionaries and in that different dictionaries have different meaning of the word sense.

Different Between Dictionaries. Different dictionaries and thesauruses will provide different divisions of words into senses. WSD accuracy is tightly coupled to the used dictionary. Research results using coarse-grained dictionaries have been much better than those using fine-grained ones. Most of the WSD research use WordNet as a reference sense inventory for English.

Amount of Samples and Semantic Knowledge Available. The amount of samples and semantic knowledge available can be solved by manually increasing them. However doing it is usually costly and undesirable. So, doing these automatically or using fewer resources is the normal way of proceeding.

2.2 *WSD Methodologies*

The WSD methods can be roughly divided into two approaches dictionary based approach and machine learning approach. For dictionary-based approach, it uses external lexical resources like Word Net, machine readable dictionary and thesaurus and so on. These two approaches are briefly discussed below;

Dictionary-Based Approach. In this approach, all the meaning of a word that needs to be disambiguated are retrieved from the dictionary. So this approach uses sense inventories or machine readable dictionary or thesaurus. It is also known as knowledge –based approach and can be divided into four types of method as bellows;[9]

LESK Algorithm. This is the first algorithm build for word sense disambiguation. This algorithm depends on dictionary and knowledge of its context at which WSD is going to perform. Basic idea of the algorithm is to find overlap among the senses of words which are to disambiguate.

Semantic Similarity. It is said that words that are related share common context and therefore the appropriate sense is chosen by those meanings found within smallest semantic distance. This semantic limitation is able to provide harmony to whole discourse; its scope has been limited to small number found in the context of target word. Various similarity measures are existing to determine how much two words are semantically related. When more than two words are there this approach also becomes extremely computationally intensive.

Selection Preference. Selectional preferences find information of the likely relations of word types, and denote common sense knowledge of classes of ideas. Simple way to study this method using measure frequency count it will find the semantic relation between words. Other method to find this kind of relation among words is conditional probability.

Heuristic Method. To forecast word sense is based on heuristics drawn from linguistic properties identified on large texts. Three types of heuristics used as a baseline for estimating WSD system are most frequent sense, one sense per discourse and one sense per collocation.

So, Knowledge-based methods have crucial advantage in oppose to corpus-based methods. Though they sometimes don't give appropriate answer and algorithms are not limited to words for which large sense data are exists and they can be easily applied to other languages if sense inventory for those languages available. A major drawback of this approach is the problem of scaling.

Machine Learning Approach. In machine learning approach system needs to be trained to perform the task of WSD. In these approaches initial inputs are target word which need to be disambiguated and the context which is portion of the text to which target word is embedded. Here the role of the classifier is to learn features and assigns senses to new unseen examples. Features are themselves provided by the words. Feature value is the occurrence of the word in the region surrounding the target word. This initial input target word is then processed after which it consist of fixed set of features to abstract information relevant to learning task. This task consist of two steps: selecting the relevant linguistic features and encoding them in a form usable in a learning algorithm. Linguistic features are divided into two classes: collocational and co-occurrence features. Here the collocation refers quantifiable position-specific relationship between two lexical items. Collocational feature encode information of specific positions located to the left or right of the target word. Typical features include the word, the root form of the word and the word's part of speech.

Part-of-Speech tagging also known as grammatical tagging is used to find the relationship with adjacent text. The techniques available based on machine learning approaches are: supervised, semi-supervised and unsupervised.

3 Techniques of Machine Learning Approach

3.1 Supervised Approaches

This approach has large number of algorithms for WSD and use machine learning techniques for disambiguation. Supervised methods require a large sense-tagged training set. Moreover in supervised learning approach, the dataset used for classification was already divided into classes. Feature of input are compared with the data available with each classes and according to highest value the input sentences will be classified in that class.[15].

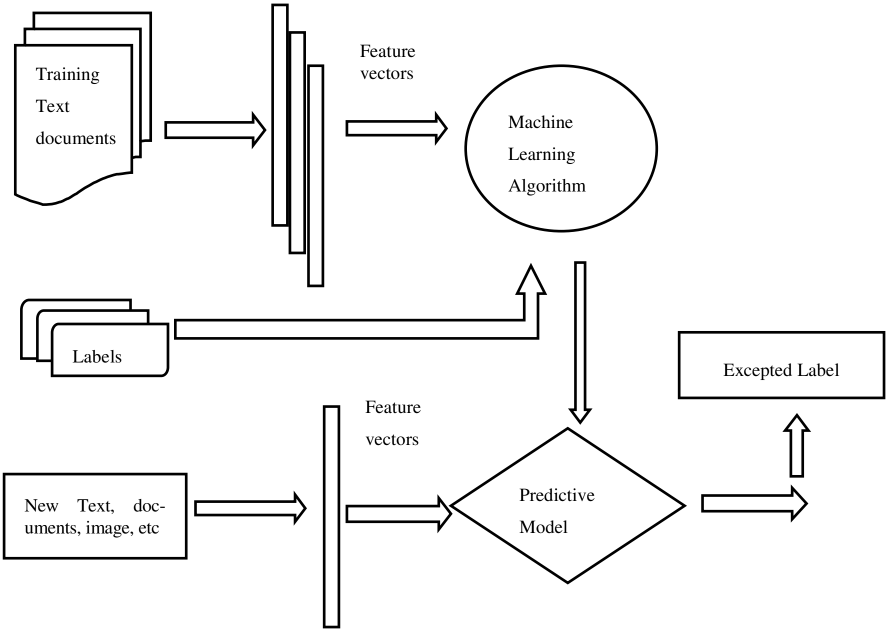

Training
Text
documents
Feature
vectors
Machine
Learning
Algorithm
Labels
Excepted Label
Feature
vectors
New Text, documents, image, etc
Predictive
Model

Fig. 2 Supervised Learning Model

In supervised learning, it is assumed that the correct (target) output values are known for each Input. So, actual output is compared with the target output, if there is a difference, an error signal should be generated by the system. This error signal helps the system to learn and reach to the desired or target output. A number of supervised approaches applied to the problem of WSD have reflected the dramatic shift from manually crafted systems to automated machine learning approach. Some of the common supervised approaches are:

Decision Tree. The set of rules are represented in the form of tree structure which recursively partition the training data set. Each internal node represents a test on feature value and each branch represents an outcome of the test and through the leaf node prediction is made. It is often regarded as a prediction tool. Some popular algorithms for learning decision trees are ID3 and C4.5. On comparison with other machine learning algorithms it was found that several supervised approaches performed better than the decision tree obtained C4.5 algorithm. If the training data is small in size decision tree suffers from prediction unreliability and decision tree also suffers from sparseness of data when there are features with large number of values. [1]

Neural Networks. Neural networks processes information based on computational model of connectionist approach. The input includes the input features and the target output. The training dataset is divided into sets which are no overlapping based on desired responses. When the network encounters new input pairs the weights are adjusted so that the output unit giving the target output has the larger activation. The network can have weights both positive and negative corresponding to correct or wrong sense choice. Neural networks are trained unless the error between the computed and the target output is minimum. Learning in neural networks is eventually updating of weights.

Decision Lists. Decision lists contains ordered set of if-then-else rules for assigning category to the test data. Features are obtained from training data which includes rules in the form of (F, S, and Score) where F represents feature value, S represents word sense. Rules are arranged in the list as per descending order of the score. For a word w represented in the form of feature vector, the winning sense for the word is the one whose feature has the maximum score in the decision list in matching the input vector. This classifier is equivalent to case statement used in programming languages. If the test returns true then the sense associated with it will be returned else next test in sequence is applied. This will continue at the end of list where default test will return the majority sense. Decision list have two senses per word.

Naïve Bayes. Naive Bayes classifier a probabilistic classifier which uses an application of Bayes's theorem. Naive Bayes is intended as a simple representative of statistical learning methods. It relies on the calculation of the conditional probability of each sense S_i of a word w given the features f_j in the context. The sense which maximizes the following formula is chosen as the most appropriate sense in context as shown in following equation;

$$\begin{aligned}\hat{S} &= \underset{S_i \in Senses_D(w)}{\operatorname{argmax}} P(S_i \mid f_1, \ldots, f_m) = \underset{S_i \in Senses_D(w)}{\operatorname{argmax}} \frac{P(f_1, \ldots, f_m \mid S_i) P(S_i)}{P(f_1, \ldots, f_m)} \\ &= \underset{S_i \in Senses_D(w)}{\operatorname{argmax}} P(S_i) \prod_{j=1}^{m} P(f_j \mid S_i),\end{aligned} \tag{1}$$

where m is the number of features, and the last formula is from the naïve assumption that the features are conditionally independent given the sense. Co-occurrence frequency in training set of sense and feature in the presence of sense is counted to evaluated the probabilities P (S_i) and P(f_j l S_i).[1]

Weakly Supervised (Bootstrapping) Methods. Bootstrapping approach rely on a small number of labeled seed instances. It works by iteratively classifying unlabeled examples and adding confidently classified examples into labeled dataset using a model learned from augmented labeled dataset in previous iteration. It can be found that the affinity information among unlabeled examples is not fully explored in this bootstrapping process. Bootstrapping is based on a local consistency assumption: examples close to labeled examples within same class will have same labels, which is also the assumption underlying many supervised learning algorithms, such as kNN. [5]

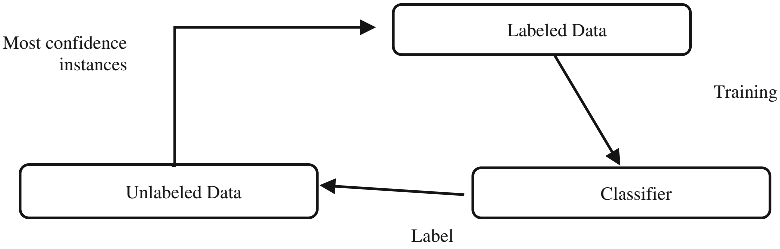


Fig. 3 Weekly supervised approach

3.2 *Semi-supervised Approaches*

An important issue of supervised method is the knowledge acquisition bottleneck. It is difficult to find the required minimum number of occurrences per each sense of word. To overcome the knowledge acquisition bottleneck faced by supervised method semi supervised method uses small annotated corpus. [11]

3.3 *Unsupervised Approaches*

This approach assumes that co-occurring words have similarities and it uses un-annotated corpus. [2]

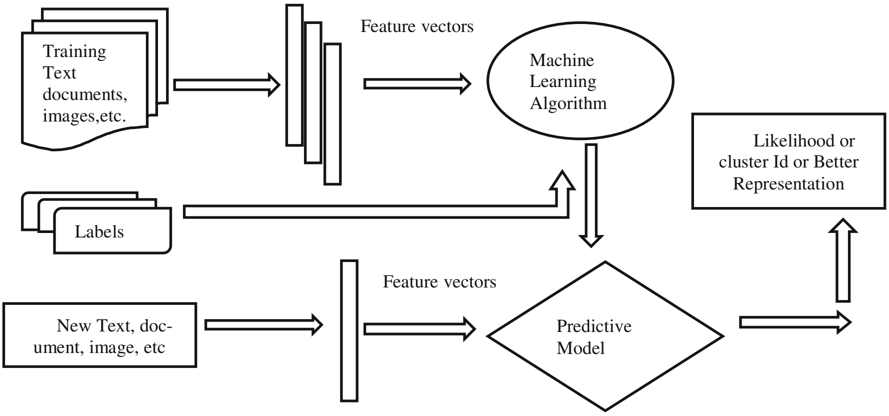

Training Text documents, images,etc.
Feature vectors
Machine Learning Algorithm
Likelihood or cluster Id or Better Representation
Labels
Feature vectors
New Text, document, image, etc
Predictive Model

Fig. 4 Unsupervised Learning Model

In machine learning, the task of unsupervised learning is to find hidden structure in unlabeled data. Corrections to the network weights are not performed by an external agent, as in many cases we also do not know what solution network should produce. Network itself has to decide what output is best for a given input and reorganizes accordingly. [15]

Context Clustering. This method is based on clustering techniques and context vector is used which is maintained for each word occurrence in the corpora. Clusters are formed for these vectors and each of such clusters corresponds to sense of word to be tested. A vector consists of every possible sense of the words. The similarity between two words m and n is provided by the cosine between the respective vectors m and n in a geometrical manner. The grouping of all the vectors gives rise to a co-occurrence matrix. After that co-occurrence matrix will be created, similarity measures will be applied after that and cosine similarity measure will going to use here from this result context vectors will be clustered and discrimination will be performed using any clustering technique from context-group discrimination and agglomerative clustering. According to this approach if a word is ambiguous then each occurrence of it is grouped to some sense cluster on the basis of similarity of context of these occurrences. Contextual similarity is measured by the cosine between two vectors corresponding between two words whose similarity is to be determined. Expectation maximization algorithm is used to perform clustering in this approach. [14]

Co-occurrence Graphs. This method will created co-occurrence graph with vertex V and edge E. where V tends to words in text and E will be added if words co-occur in their relation according to syntax in same paragraph or in text. If the target word is given then graph will be created and for graph adjacency matrix will be created then Markov clustering method will be applied to find meaning of word. For given target word local graph G is created and for that adjacency matrix is created then Markov clustering algorithm is applied to get sense of word. [14]

Word Clustering. This technique is similar to context clustering in terms finding sense but it will cluster those words which are semantically identical and so it gives a related meaning of word. In this technique words having similar meanings are assigned to the same cluster. The similarity between the words is given by syntactical dependency. If W consist of words which are similar to w_m then a tree is formed initially with only one node w_m and a node w_i will have a child node w_m when w_i is found to be the word with most similar meaning to w_m. [14]

4 Evaluation and Result

To evaluate the Myanmar WSD system, we need to measure the accuracy and performance of the machine learning techniques. So we need to calculate coverage, precision and recall from the results. [10]

Coverage C- Coverage is the measure of percentage of words in the test data for which the WSD system has given sense assignment. It is represented as:

$$C= \text{Answers provided/Total answers to provide} \tag{2}$$

Precision P- Precision is the measure of ratio of correct answers provided to the answers provided. It is represented as:

$$P= \text{Correct answers provided/Answers provided} \tag{3}$$

Recall R- Recall is the measure of the ratio of correct answers provided to the total number of answers to provide. It is represented as:

$$R= \text{Correct answers provided/ Total answers to provide} \tag{4}$$

F1 measure- F1 measure is the weighted harmonic mean of precision and recall. It is represented as:

$$\text{F1 measure}= (2 * P * R)/ (P+R) \tag{5}$$

Precision is found to be equal to Recall when coverage is 100%.

Table 1 Comparative Study of four supervised classifier

1. Approach	**2. Average Precision**	**3. Average Accuracy**
4. Decision Tree	5.64.13%	6.60.9%
7. Neural Network	8.63.3%	9.64.5%
10. Decision List	11.96%	12.63.9%
13.Naïve Bayes	14.97%	15.58.32%

In the above table a comparison of different supervised approaches has been given based on their Average precision and Average baseline accuracy.

4.1 Comparison [9]

Comparing and evaluating different WSD systems is very difficult, because of the different test sets, sense inventories or machine readable dictionaries, and knowledge resources required. From this research review of WSD approaches, Knowledge-based approach is depend on sense inventories or machine readable dictionaries which are tend to tether with their own interpretation of word, if word which is not available in dictionary than it will be difficult to find its meaning. Supervised approach gives better result than others but it is heavily depends on manually created sense-annotated text and to manually create these sense-annotated text requires lots of time and effort. Unsupervised algorithm is not depends on any sense-annotated data or machine readable dictionaries. It finds the sense on assumption that words that occur frequently together have similar meaning. Comparison of these described three approaches is shown in below Table 2.

Table 2 Comparison of WSD Approaches

16. Approach	17. Advantage	18. Disadvantage
19. Knowledge-based (or) dictionary-based	20. Gives higher precision	21. Overlap based algorithms, suffers from overlap sparsity, dictionary definitions
22. Supervised 23. Approach	24. Better than others	25. Require large sense annotated data not for resource scarce language
26. Unsupervised Approach	27. No need of any sense inventory and sense annotated corpora	28. Difficult to implement, performance is always inferior to that of other two approaches.

5 Conclusion and Future Work

In this paper we discussed about the machine learning techniques for word sense disambiguation in natural language processing. Also describes the most well-known classification algorithms as supervised learning and clustering algorithms as unsupervised learning and dictionary-based approach for word sense disambiguation. The aim behind this study was to learn their key ideas. Both supervised and unsupervised methods have advantages and disadvantages: on one hand, it is possible to apply simple supervised methods to disambiguate a small pre-defined set of words. Whereas, for more robust applications, unsupervised methods seems to be more suitable as they can deal with a bigger portion of the lexicon. My future work is mining opinions on social network comments of Myanmar language to detect negative comments and then applying WSD on that to find correct sense on word.

References

1. Boshra, F., Al-Bayaty, Z., Josh, S: Comparative Analysis between Navïe Bayes Algorithm and Decision Tree to Solve WSD Using Empirical Approach
2. Chaplot, D.S.: Literature Survey on Unsupervised Word Sense Disambiguation
3. Tatar, D.: Word Sense Disambiguation By Machine Learning Approach: A Short Survey
4. Escudero, G., Màrquez, L., Rigau1, G.: Naive Bayes and Exemplar-Based approaches to Word Sense Disambiguation Revisited
5. Escudero, G., Mhrquez, L., Rigau, G.: A Comparison between Supervised Learning Algorithms for Word Sense Disambiguation
6. Kaur, J.: Word Sense Disambiguation (WSD)
7. Zampieri, M.: A Supervised Machine Learning Method for Word Sense Disambiguation of Portuguese Nouns
8. Sasaki, M.: Latent Semantic Word Sense Disambiguation Using Global Co-Occurrence Information
9. Patel, N., Patel, B., Parikh, R., Bhatt, B.: A Survey: Word Sense Disambiguation
10. Borah, P.P., Talukdar, G., Baruah, A : Approaches for Word Sense Disambiguation – A Survey
11. Mante, R., Kshirsagar, M., Chatur, P.: A Review Of Literature On Word Sense Disambiguation
12. Jain, R., Vidyapith, B., Rajasthan, B.V.: Sense Disambiguation Techniques: A Survey
13. Shallu, S.: A Survey of Word-sense Disambiguation Effective Techniques and Methods for Indian Languages. shallu.2.chd@gmail.com
14. Steve Lawrence Manion: Unsupervised Knowledge-Based Word Sense Disambiguation: Exploration & Evaluation of Semantic
15. Rawat, S.: Word Sense Disambiguation and Classification Algorithms: A Review

Part III

Advance Image and Signal Processing Technology and Applications

A Graph Representative Structure for Detecting Automorphic Graphs

Yu Wai Hlaing and Kyaw May Oo

Abstract Graphs are prevalently used to model the relationships between objects in various domains. Storing the graphs into large databases is a challenging task as it deals with efficient space and time management. Unlike item sets in huge transactional databases, it becomes essential to ensure the consistency of graph databases since relationships among edges of a graph are predominant. One of the necessary procedures required is a mechanism to check whether two graphs are automorphic(duplicated) or not. Difficulty in identifying and eliminating the automorphic graphs is a challenging problem to the research community. In this paper, we propose a graph representative structure that is called graph code. There are three main phases: preprocessing, code generation and code matching. In preprocessing phase, vertex list, edge list and adjacent edge information are generated for input graph. In code generation, edge dictionary plays an important role. The edge dictionary and adjacent edge information are used to generate graph codes. In code matching, the new graph code is compared with those of other graphs in graph dataset to determine whether they are automorphic or not. The experimental results and comparisons offer a positive response to the proposed structure.

Keywords Edge dictionary · Graph automorphism · Graph code · Graph representative structure

1 Introduction

A graph describes relationships over a set of entities. With node and edge labels, a graph can describe the attributes of both the entity set and the relation.

Y.W. Hlaing(✉)
University of Computer Studies, Yangon, Myanmar
e-mail: yuwaihlaing.1987@gmail.com

K.M. Oo
University of Information Technology, Yangon, Myanmar
e-mail: kmayoo19@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_19

Labeled graphs appear in many research domains such as drug design [1], protein structure comparison [2], video indexing [3], and web information [4]. In addition, rapidly increasing Web sites and XML documents can also be modeled as graphs. Therefore, it is evident that graph data will become more and more prevalently used in the near future.

Various conferences over the past few years on mining graphs have motivated researchers to focus on the importance of mining graph data. One of the major perceptions concerned in graph mining is discovering frequent patterns [5], [6], [7], and [8]. The key operation required by any frequent subgraph discovery algorithm is a mechanism to check whether two subgraphs are identical or not. Many graph databases such as chemical graphs have more than one vertex with the same label. There are possibilities that the same graph is stored more than once in the graph database leading to adverse results of mining. So, there is a need to represent graphs efficiently in order to identify automorphic graphs.

Structures that can be represented as graphs are based on graph theory [9]. Graph databases apply graph theory to store information about the relationships between entities in terms of graphs. There are many different structures that can be represented as graphs. Perhaps the simplest graph representation of a graph is as an unordered edge sequences. Each edge contains a pair of node indices and, possibly, associated information such as an edge weight. Adjacency array support easy access to the edges leaving any particular node, we can store the edges leaving any node in an array. If no additional information is stored with the edges, this array will just contain the indices of the target nodes. Next one is adjacency list. Size of the array is equal to number of vertices. Let the array be array []. An entry array[i] represents the linked list of vertices adjacent to the i^{th} vertex. An *n*-node graph can also be represented by an n×n adjacency matrix *A*. $A_{i,j}$ is 1 if (i, j) $\in$ E and 0 otherwise. Among them the two most popular graph representative structures are adjacency list and adjacency matrix.

The most important one is to be compact and less time needed in generating representative structures for graph. Our graph code is a new way of representing graph data to process graph queries without verification between graph structures. It is developed to process labeled, undirected chemical compound graphs in the area of chemical informatics.

The rest of the paper is organized as follows. Section 2 reviews the related works in this area. Section 3 presents the formal definitions and notations used for the proposed research work. Section 4 explains proposed graph representative structure. Section 5 shows the experimental results of proposed structure using AIDS Antiviral Screen Compound Dataset. In section 6 concludes our paper.

2 Related Works

Over the years, a number of different structures have been developed to represent graphs more and more efficiently. Developing such structures is particularly challenging in terms of storage space and computational time.

Michihiro Kuramochi et al. proposed a canonical labeling [10] for representing graphs. Canonical label of a graph is nothing more than a code that uniquely identifies the graph such that if two graphs are isomorphic to each other, they will be assigned the same code. Canonical label of a graph is as the string obtained by concatenating the upper triangular entries of the graph's adjacency matrix. Once the canonical label has been obtained, the adjacency matrix representation is discarded. The disadvantage of this structure is if a graph has |V| vertices, the complexity of determining its canonical label is in $O(|V!|)$ making it impractical even for modern size graphs.

The search space of canonical labeling can be reduced with vertex invariants. Vertex invariant is a well-known technique in which we can partition the vertices by their degrees and labels. Then, we try all the possible permutations of vertices inside each partition. Vertex invariants do not asymptotically change the computational complexity of canonical labeling. For example, if a given graph is regular, we cannot create fine partitions and vertex invariants do not reduce the search space [11].

R Vijayalakshmi et al. [12] propose F-GAF: a novel approach for detection and elimination of automorphic graphs in graph database. F-GAF uses edge-based graph representation. It involves three main phases: preprocessing, feature extraction and pattern matching. In preprocessing, edge list of the input graph is generated. In feature extraction phase, grid code is generated. In pattern matching the new grid code is compared with those of other graphs in graph database. The weak points of this structure are it requires exhausted enumeration to generate grid code and it needs large number of comparisons to determine automorphic graphs.

3 Definitions and Notations

This section presents the key concepts, notations and terminology used in this paper, which include: labeled graphs, graph isomorphism, graph automorphism, and graph code representation.

Definition 1 (Labeled Graph). A labeled graph is a four-element tuple $G = (V, E, \Sigma, f)$where V is a set of vertices and E is a set undirected edges joining two distinct vertices. Σ is the set of vertex and edge labels and f maps vertices and edges to their labels.

Definition 2 (Graph Isomorphism). An isomorphism of graphs G and H is a bijection between the vertex sets of G and H; $f: V(G) \rightarrow V(H)$such that any two vertices u and v of G are adjacent in G if and only if $f(u)$ and $f(v)$ are adjacent in H.

Definition 3 (Graph Automorphism). An automorphism of a graph $G = (V, E, \Sigma, f)$ is a permutation σ of the vertex set V, such that the pair of vertices (u, v) form an edge if and only if the pair $(\sigma(u), \sigma(v))$ also form an edge. That is, it is a graph isomorphism from G to itself.

Definition 4 (Graph Code Representation). For a graph G, the code of G, denoted by $c(G)$ is in the form $e_{id}\{(v), eid_{adj}, \ldots\}\ldots$ depending on adjacent edge information of preprocessing. e_{id} is the edge id, v is vertex label on which two edges are connected, eid_{adj} is list of adjacent edge ids for this edge.

An example labeled graph is shown in Fig. 1.

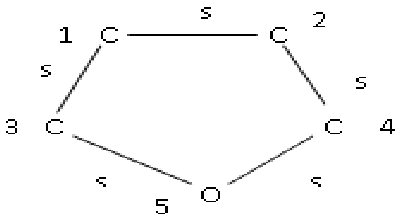

1
C
s
C
2
s
s
3
C
C
4
s
s
5
O

Fig. 1 A Labeled Graph (G)

An example graph code of G is $c(G)$=1{(c)1,1} 1{(c)1,2} 1{(c)1,2} 2{(c)1,(0)2} 2{(c)1,(0)2}.

4 Our Graph Representative Structure

In this section, our proposed graph representative structure is described that applied on a chemical compound graph dataset for efficient detection of automorphic chemical graphs. There are three main phases to generate graph code. The first phase is preprocessing. And the other is code generation. Edge Dictionary plays an important role in generating graph code. The final phase is code matching phase. The overall structure is described in Fig. 2.

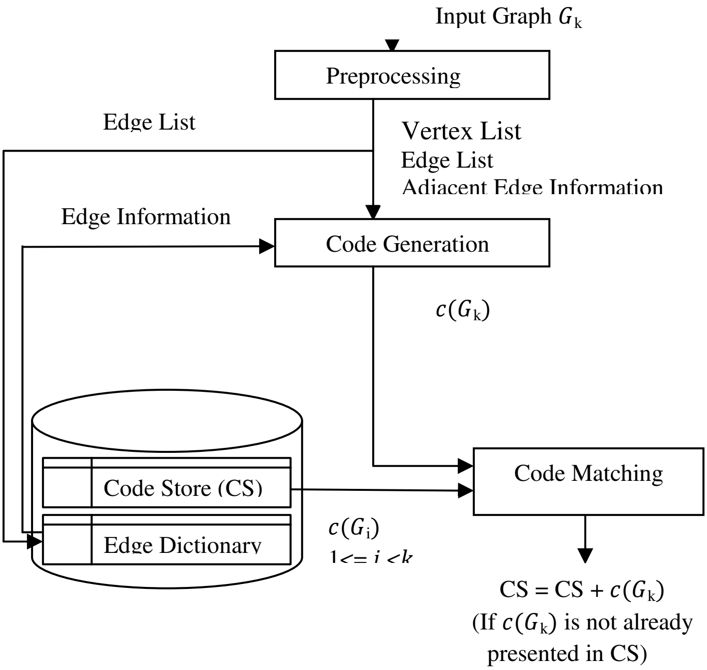

Input Graph G_k
Preprocessing
Edge List
Vertex List
Edge List
Adjacent Edge Information
Edge Information
Code Generation
$c(G_k)$
Code Store (CS)
Edge Dictionary
Code Matching
$c(G_i)$
$1<= i <k$
CS = CS + $c(G_k)$
(If $c(G_k)$ is not already presented in CS)

Fig. 2 Overall Structure of Proposed Work

4.1 Preprocessing

In this step, for a graph the vertex list, edge list, and adjacent edge information are generated. Each vertex in the graph is assigned with unique id.

Vertex id :	1	2	3	4	5
Vertex List :	C	C	C	C	O

Then the edge list of the graph is defined as (V_{id},L,V_{id}) where V_{id} is the vertex id, L is the edge label. The output of this step is as follows.

V_{id},L,V_{id} :	<1,s,2>	<1,s,3>	<2,s,4>	<3,s,5>	<4,s,5>
Edge List:	<C,s,C>	<C,s,C>	<C,s,C>	<C,s,O>	<C,s,O>

4.2 Edge Dictionary

When a graph introduced to the database, the edges in the graph are added into edge dictionary if these edges are not already existed in edge dictionary. In this dictionary, these edges are assigned with global unique ids for further processing. Fig. 3 shows the example edge dictionary for input graphs.

Id	Edge
1	<C,s,C>
2	<C,s,O>
3	<C,s,C>
...	…

Fig. 3 Edge Dictionary

4.3 Code Generation

A graph is represented holistically into a graph code that captures the structural representation of the graph. Every edge in the graph is assigned with global unique identifier already defined in the edge dictionary. Instead of using the edge itself, using the edge id in the dictionary can have advantages in three ways:

- Firstly, using the edge id in the code saves the amount of storage space.
- Secondly, using the same id for the duplicated edge is effective when constructing the edge code.
- Thirdly, using the edge id in the code reduces the time of matching graphs.

Most of the chemical graphs have a lot of common edges. So, edge dictionary uses a little memory space. Edge dictionary and adjacent edge information are used to generate graph code. Graph code of the input graph is stored in code store (CS) if it is not already existed in CS.

4.4 Code Matching

After generating the graph code of the k^{th} graph, Code of G_{k} is compared with each graph code of G_{i}, $1 \leq i < k$ in code store (CS). If the graph code of G_{k} is not the same code as one of G_{i}, concludes that the graph is not automorphic graph and its graph code is stored in CS. Otherwise, the graphs are automorphic and then terminate the process. In this process of comparison in the specified order, if any of these codes are different, the algorithm immediately concludes that the graphs are different.

5 Experimental Results

The experiment described in this section use the AIDS Antiviral Screen Dataset. The AIDS Antiviral Screen Dataset from Development Therapeutics Program NCI/NIH is available publicly. The dataset contains 43,906 chemical compounds. Each compound has 32 vertices and 34 edges in average. The maximum one has 188 vertices and 196 edges. In this section, we evaluate the storage space of our graph codes with the storage space required by other two formats: xml file and image file (.png). In Fig. 4, it can be seen that the storage space of our graph codes is significantly less than the other two formats.

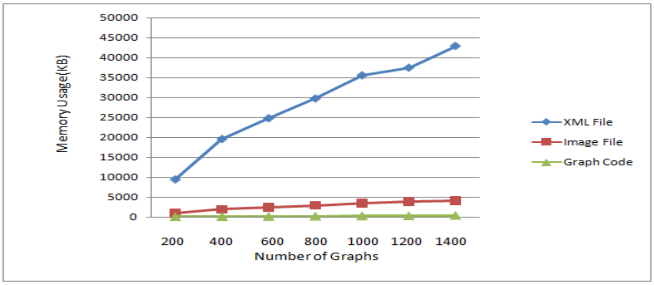


Fig. 4 Analysis of Storage Space for Various Numbers of Graphs between Graph Code, XML File, Image File

The time complexity is commonly estimated by counting the number of elementary operations performed by an algorithm where an elementary operation takes a fixed amount of time to perform. In proposed approach, computational time complexity for generating graph code takes $O(NE + (E * (E - 1))/2)$ where E is the number of edges, V is the number of vertices and N is the size of edge dictionary. Proposed approach takes NE comparisons is required to check edge dictionary whether the edges in the input graph already exist in it or not. For getting adjacent edge list, $(E * (E - 1))/2$ comparisons are required. Table.1 describes the computational time complexity between three techniques for generating graph representative structure. Tabel.2 shows the analysis of computational time complexity for code generation between three techniques. Graph code can reduce at least 10^4 times computational complexity when compared to canonical labeling. But it consumes a little more time than F-GAF for code generation.

Table 1 Computational Time Complexity for Code Generation between Three Techniques

Technique	Computational Complexity
Canonical Labeling	$O(V!)$
F-GAF	$O(4E)$
Graph Code	$O(NE + (E * (E - 1))/2)$

Table 2 Analysis of Computational Time Complexity for Code Generation between Three Techniques

No. of vertices	No. of edges	Canonical Labeling	F-GAF	GraphCode(N=19)
10	10	3.63×10^{6}	4.0×10	1.45×10^{2}
20	20	2.43×10^{18}	8.0×10	5.90×10^{2}
40	40	8.16×10^{47}	1.60×10^{2}	2.38×10^{3}
60	60	8.32×10^{81}	2.40×10^{2}	5.37×10^{3}
80	80	7.16×10^{118}	3.20×10^{2}	9.56×10^{3}

Table. 3 shows the analysis of computational time complexity for detecting automorphic graphs between canonical labeling, F-GAF and proposed approach. The analysis of comparisons for detecting automorphic graphs required by three techniques is shown in table. 4. Proposed approach can reduce at least 10^2 times computational complexity when compared to canonical labeling. It can lessen at least 1 time computational time complexity than F-GAF.

Table 3 Computational Time Complexity for Detecting Automorphic Graphs between Three Techniques

Technique	Computational Complexity
Canonical Labeling	$O\ ((V!)\ +\ k\ *\ ((V2) - V)/2))$
F-GAF	$O\ ((4E)\ +\ k\ *\ (2 + 3E + (5V)^{2} - V))$
Graph Code	$O\ ((NE + (E * (E - 1))/2)\ +\ k\ *\ 2E * (V - 2))$

Where,
V = number of vertices
E = number of edges
N = size of edge dictionary
k = number of graphs

Table 4 Analysis of Computational Time Complexity for Detecting Automorphic Graphs between Three Techniques in 100 Graphs

No. of vertices	No. of edges	Canonical Labeling	F-GAF	GraphCode(N=19)
10	10	$3.63 \text{x}10^{6}$	$2.52 \text{x}10^{5}$	$1.61 \text{x}10^{4}$
20	20	$2.43 \text{x}10^{18}$	$1.00 \text{x}10^{6}$	$7.26 \text{x}10^{4}$
40	40	$8.16 \text{x}10^{47}$	$4.01 \text{x}10^{6}$	$3.06 \text{x}10^{5}$
60	60	$8.32 \text{x}10^{81}$	$9.01 \text{x}10^{6}$	$7.01 \text{x}10^{5}$
80	80	$7.16 \text{x}10^{118}$	$1.60 \text{x}10^{7}$	$1.26 \text{x}10^{6}$

6 Conclusion and Future Work

We propose a new way of representing a graph holistically via its graph code. The edge dictionary is efficiently used to narrow down the search space of graph code. Automorphic graphs can be detected by only comparing graph codes instead of expensive automorphism testing. From our experimental results, our graph code outperforms the existing storage structures and methods. Efficient graph querying using graph codes is going to be observed as a future work.

References

1. Borgelt, C., Berthold, M.R.: Mining molecular fragments: finding relevant substructures of molecues. In: ICDM, pp. 51–58 (2002)
2. Huan, J., Wang, W., Bandyopadhyay, D., Snoeyink, J., Prins, J., Tropsha, A.: Mining protein family specific residue packing patterns from protein structure graphs. In: Proceedings of the 8th Annual International Conference on Research in Computational Molecular Biology (RECOMB), pp. 308–315 (2004)
3. Messmer, B.T., Bunke, H.: A Decision Tree Approach to Graph and Subgraph Isomorphism Detection. Pattern Recoginition **32**(12), 1979–1998 (1991)
4. Raghavan, S., Garcia-Molina, H.: Representing web graphs. In: Proceeding of IEEE International Conference on Data Engineering (2003)
5. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceeding of the 11th International Conference on Data Engineering (ICDE), pp 3–14. IEEE Press (1995)
6. Agrawal, R., srikant, R.: Fast algorithms for mining associations rules. In: Proceeding of 20th International Conference on Very Large Databases (VLDB), pp. 487–499, September 1994
7. Chen, C.W.K., Yun, D.Y.Y.: Unifying graph-matching problem with a practical solution. In: Proceeding of International Conference on Systems, Signals, Control, Computers, September 1998

8. Chittimoori, R.N., Holder, L.B., Cook, D.J.: Applying the SUBDUE structure discovery system to the chemical toxicity domain. In: Proceeding of the 12th International Florida AI Research Society Conference, pp. 90–94 (1999)
9. Bondy, J.A., Murty, U.S.R.: Graph Theory With Applications. Elsevier Science Publishing Co. Inc., New York (1976)
10. Kuramochi, M., Karypis, G.: An efficient algorithm for discoverying frequent subgraphs. In: Proc. 2001 International Conference on Data Mining (ICDM 2001), San Jose, CA, pp. 313–320, November 2001
11. Fortin, S.: The graph isomorphism problem. Technical Report TR 96-20, Department of Computing Science, University of Alberta (1996)
12. Vijayalakshmi, R., Nadarajan, R., Nirmala, P., Thilaga, M.: A Novel Approach for Detection and Elimination of Automorphic Graphs in Graph Database. Int. J. Open Problems Compt. Math. **3**(1), March 2010

A Method of Wavelength Selection of Multispectral Image for the Separation of Tumor Region

Kayo Yoshimoto, Hideya Takahashi, Kenji Yamada and Shigeto Yoshida

Abstract Endoscopy is widely used for examinations and diagnosis of cancer. However, it is difficult to accurately identify the tumor region from the endoscopic image unless experienced experts. In this paper, we propose a method of the separation of tumor region for endoscope surgery and examination by using multispectral images. By selecting the effective wavelength to identify from multispectral images, the tumor region can be separate. We confirm the proposed method by several test targets. The proposed method would help for endoscopic surgery.

Keywords Endoscopy · Multispectral imaging · Principal component analysis (PCA) · Wavelength selection

1 Introduction

Minimally invasive surgical techniques for endoscope become widely used, for example, laparoscopic operation, EMR (Endoscopic Mucosal Resection), ESD (Endoscopic Submucosal Dissection), robotic surgery and so on. These surgery can

---

K. Yoshimoto(✉) · H. Takahashi
Graduate School of Engineering, Osaka City University, 3-3-138 Sugimoto,
Sumiyoshi-ku, Osaka 558-8585, Japan
e-mail: {yoshimoto,hideya}@elec.eng.osaka-cu.ac.jp

K. Yamada
Graduate School of Medicine, Osaka University, 2-2 Yamadaoka,
Suita, Osaka 565-0871, Japan
e-mail: k-yamada@sahs.med.osaka-u.ac.jp

S. Yoshida
Department of Endoscopy, Hiroshima University Hospital, 1-2-3 Kasumi,
Minami-ku, Hiroshima 734-8551, Japan
e-mail: yoshida7@hiroshima-u.ac.jp

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_20

reduce the burden on the patient and recovery times. To improve the possibility of cure, early detection of cancer is very important. If cancer has spread the mucosal layer but not infiltrate deep into the submucosa, the treatment can finish only by endoscopic treatment without abdominal operation. For early detection of cancer, the endoscope high-resolution image can be obtained or endoscope for magnifying observation was developed. On the other hand, in operation for cancer, it is necessary to remove the cancer cells completely and minimum. This is for reducing the risk of cancer recurrence and the burden on the patient. In the other words, a particular tumor region is very important. The visualization of cancer could support the examiner during endoscopy and has been actively researched[1]–[9]. For example, chromoendoscopy[1] and fluorescence imaging[2] use external agents such as dyes to highlight specific features of the mucosa. Narrow band imaging (NBI)[3][4] and optimal band imaging(OBI)[5] are a method of highlighting the micro structure and vascular network structure of the mucosal surface. These are based that the cancer and the other have different the absorption characteristics for wavelengths[2][6]. The mortality rates of the stomach cancer in Japan, may have declined by the introduction of these diagnostic method[10]. However, even if it possible to observe the structure and the vascular pattern by these methods, it is difficult to accurately identify the tumor region from the endoscopic image unless experienced experts.

In this paper, we propose a method for specifying tumor region by using the difference in the spectra of light diffused by tumor and normal tissue. Since the characteristics depend on the wavelength, the difference between tumor and normal tissue emphasized wavelength is automatically selected and only the narrow-band image at that wavelength is used for the separation in the proposed method. Focusing on the second-order differential absorbance as a feature, the optimum wavelength is automatically selected by using a principal component analysis. Furthermore, it is verified whether there is a significant difference in the tumor region and the normal region in selected wavelength. This paper is organized as follows. In section 2, we describe the proposed method. In Section 3, the proposed system is evaluated by using a cooled CCD camera and a Liquid Crystal Tunable Filter (LCTF) and ex-vivo sample and discuss about the experimental results. Finally, we give some conclusions in Section 4.

2 Method for Extracting Tumor Region Using Multispectral Image

2.1 Absorption Characteristics of Biological Tissue

The cancer cells grow by absorbing nutrition and oxygen from blood vessel around them and make new blood vessel. In other words, blood vessels is an essential element for the cancer. The tumor has may blood vessels and the tumor vessels have different characteristic structures of normal blood vessels.

On the other hand, generally, amount of the light in short-wavelength cannot reach deep into the biological tissue and it is observed as the reflected light because it absorbs and scatters in the surface. The longer wavelength, light deeply propagation into the biological tissue. A part of the light can be observed as reflected light, the other is transmitted light. Such reaching depth of wavelength dependence caused by the specific absorption of the blood hemoglobin and light scattering property of biological tissue. Therefore, by narrowing the bandwidth of the wavelength of light, the features of the biological tissue at each wavelength can be captured. The multispectral imaging is a technique of measuring a discrete spectrum with only a few wavelengths. As in the previous discussion, the features of tumor and normal vessels is difference at a certain wavelength. So, we can extract the tumor region by the multispectral imaging using only the narrow band image at the wavelength that emphasize those difference. In this paper, we proposed a method for extracting tumor region by using the difference of the absorbance between the tumor region and the normal one. The absorbance is defined as the ratio of the transmitted light intensity to the incident light intensity at a specified wavelength based on the Lambert-Beer law. The absorbance $A(\lambda)$ at wavelength λ is given by

$$A(\lambda) = \log_{10}(I_s(\lambda)/I(\lambda)) \tag{1}$$

where $I_s(\lambda)$ is the light intensity from standard white board, $I(\lambda)$ is the light intensity from target at wavelength λ.

Then, to emphasize small differences in absorbance spectra, the absorbance of the target is processed by secondary differentiation. The secondary differentiation of the absorbance $d^2A(\lambda)$ is given by

$$d^2A(\lambda) = A(\lambda - \Delta\lambda) - 2A(\lambda) + A(\lambda + \Delta\lambda) \tag{2}$$

where $\Delta\lambda$ is the sampling interval of the wavelength. This differentiation process also has the effect of removing the additive baseline fluctuation generated by the difference in the surface conditions of the target.

2.2 Selection of Optimum Wavelength

The biological tissue including tumors and cancers can be classified into the following four categories: blood vessel and mucosa in tumor region, blood vessel and mucosa in normal region. The optimum wavelength for the tumor segmentation is selected under three conditions:

1: The difference of the absorbance secondary differential value between tumor and normal regions is large.
2: The difference of the absorbance secondary differential value between blood vessel and mucosa in tumor regions is small.
3: The difference of the absorbance secondary differential value between blood vessel and mucosa in normal regions is small.

In this study, we proposed a method for automatically selecting optimum combination of some wavelength by using principal component analysis(PCA) as a multivariate analysis.

PCA is the most typical analysis in multivariate analysis used for dimension reduction of the high-dimensional datasets. PCA is a procedure that uses an orthogonal transformation to convert a mutually correlative variables set into an unrelated variables set called principal component. This conversion is geometrically equivalent to the coordinate rotation and translation in a multidimensional vector space, and is algebraically equivalent to obtain an eigenvalues of the covariance matrix or correlation matrix of the variables. The transformation of the PCA is defined as the first principal component has the largest possible variance. Subsequent component has the highest variance under the constraint that is orthogonal to the preceding components.

Then, let us consider explanatory variables for PCA. Based on the above conditions, the difference of the absorbance secondary differential value between blood vessel in tumor and blood vessel in normal regions $\boldsymbol{x}$, that between blood vessel and mucosa in tumor regions $\boldsymbol{y}$, and, that in normal region $\boldsymbol{z}$ are used for explanatory variables. Here, we focused on the difference between the blood vessel as a sample to measure the difference between the tumor and normal region. These variables are given by

$$x(\lambda_i) = |tv(\lambda_i) - nv(\lambda_i)| \; (i = 1, 2, ..., M) \tag{3}$$

$$y(\lambda_i) = |tv(\lambda_i) - tm(\lambda_i)| \; (i = 1, 2, ..., M) \tag{4}$$

$$z(\lambda_i) = |nv(\lambda_i) - nm(\lambda_i)| \; (i = 1, 2, ..., M) \tag{5}$$

where M is the number of measured wavelengths, $tv(\lambda_i)$, $tm(\lambda_i)$, $nv(\lambda_i)$ and $nm(\lambda_i)$ are the absorbance secondary differential value of blood vessel and mucosa in tumor regions, and these in normal one at wavelength λ_i, respectively. These values $tv(\lambda_i), tm(\lambda_i), nv(\lambda_i), nm(\lambda_i)$ are given as a typical example by the preliminary sampling at each wavelength. The k-th principal component score $Z_k(\lambda_i)$ (k=1,2,3, i=1,...M) can be expressed by using the eigenvector $\boldsymbol{v}_k = [v_{k1}, v_{k2}, v_{k3}]$ correspond to the k-the principal component as follows:

$$Z_k(\lambda_i) = v_{k1}(x(\lambda_i) - \bar{x}) + v_{k2}(y(\lambda_i) - \bar{y}) + v_{k3}(z(\lambda_i) - \bar{z}) \tag{6}$$

Here, the sign of the weight of each term in the k-th principal component score, that is the eigenvector, represent the properties of each variable. In this study, the optimum wavelength is apply the above conditions. For this reason, the wavelength should be selected in the order of the principal component score in the principal component so as the sigh of eigenvector to be successively positive, negative, and negative, or, in the order from the lowest selected principal component score in the principal component so as the sigh of eigenvector to be successively negative, positive, and positive.

Then, we confirm whether the wavelength selected by the PCA is the optimal wavelength for segmentation of tumor region. The verification is carried out based on statistical hypothesis testing. The test is used to find out if there is a real difference

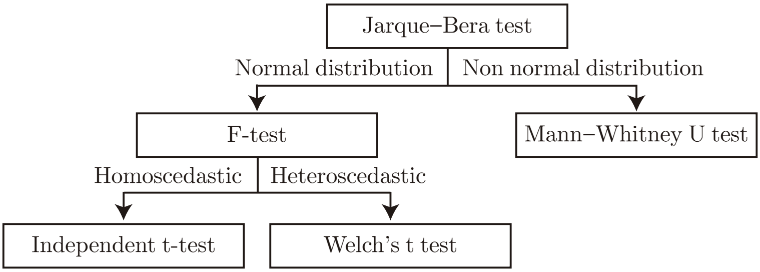


Fig. 1 Flow chart for selecting test method

between a typical example of the tumor region and the normal region at selected wavelength by the preliminary sampling. That is, the test is performed in the following four patterns:

1: The blood vessel in tumor region *vs*. the blood vessel in normal region.
2: The blood vessel in tumor region *vs*. the mucosa in normal region.
3: The mucosa in tumor region *vs*. the blood vessel in normal region.
4: The mucosa in tumor region *vs*. the mucosa in normal region.

If a significant difference is not found even one, the next wavelength is selected, and the test is performed similarly. Only the wavelength that significance is observed is selected. The test is selected by the flow chart as shown in Fig. 1.

2.3 Extraction of the Tumor Region Using the Correlation Method

Then, to separate the tumor region, the absorbancethe correlation matching between each pixel of captured image and the average absorbance of the tumor vessel is analyzed at selected wavelengths. Here, the absorbance of the blood vessel in tumor region is used as a reference. Because, the wavelength is selected so that the difference in absorbance of the tumor region and the normal region is large and the difference in absorbance of the blood vessel and the mucosa in tumor region is small. The correlation coefficient R is expressed by the following equation:

$$R = \frac{\sum_{i=1}^{L} a(\lambda_i)x(\lambda_i)}{\sqrt{\sum_{i=1}^{L} a^2(\lambda_i)}\sqrt{\sum_{i=1}^{L} x^2(\lambda_i)}} \tag{7}$$

where, $a(\lambda_i)$ is the absorbance secondary differential value of the tumor vessel, $x(\lambda_i)$ is the absorbance secondary differential value of any pixel in the obtained image at selected wavelengths $\lambda_1, \lambda_2, ..., \lambda_L$, and L is the number of selected wavelengths.

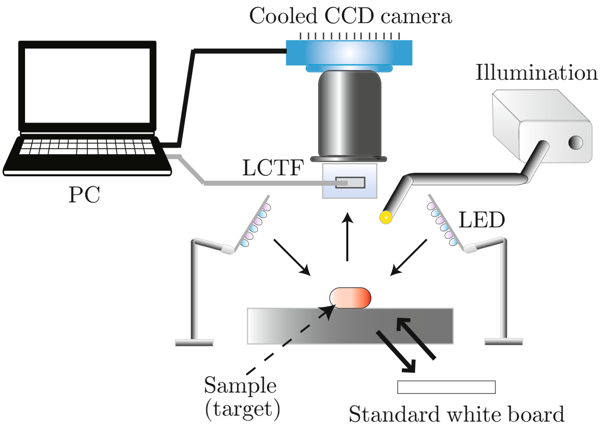


Fig. 2 Experimental setup

Then, to enhance the correlation value, we use the following equation:

$$\hat{R} = \frac{\sum_{i=1}^{L}(a(\lambda_i) - g(\lambda_i))(a(\lambda_i) - g(\lambda_i))}{\sqrt{\sum_{i=1}^{L}(a(\lambda_i) - g(\lambda_i))^2}\sqrt{\sum_{i=1}^{L}(x(\lambda_i) - g(\lambda_i))^2}} \tag{8}$$

$$g(\lambda_i) = \frac{a(\lambda_i) + x(\lambda_i)}{3}(i = 1, 2, ..., L) \tag{9}$$

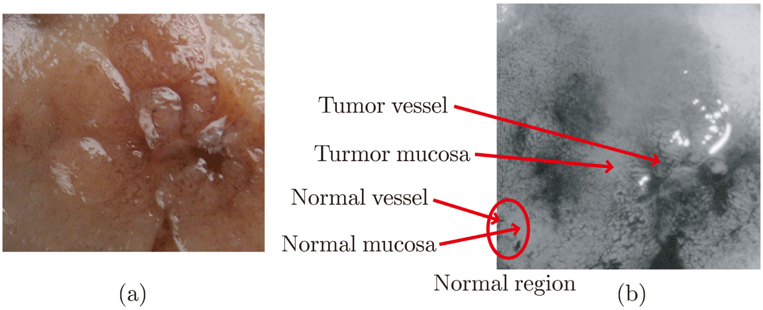
where, $g(\lambda_i)$ is the center of gravity of vectors $\boldsymbol{X}(x(\lambda_1), x(\lambda_2), ..., x(\lambda_i))$, $\boldsymbol{A}(a(\lambda_1), a(\lambda_2), ..., a(\lambda_i))$,and the origin $\boldsymbol{O}$.

3 Experiments

3.1 Experimental Setup

A schematic diagram of the experimental setup is shown in Fig. 2. Multispectral images are obtained by a cooled CCD camera(Apogee Co. Ltd.) and a LCTF(Cambridge Research Inc.) which can scan its central wavelength from 400 nm to 720 nm. Table 1 and Table 2 show specifications of the LCTF and the CCD respectively. We took a total of 65 narrow band images at 5nm intervals in the range from 400nm to 720nm and obtained spectral information from them.

At first, we obtained typical examples of the absorbance of the tumor vessel, the tumor mucosa, the normal vessel, and the normal mucosa. We selected 20 samples in



(a) (b)

Fig. 3 Target object. (a) stomach polyp, (b)narrow band image at 540nm

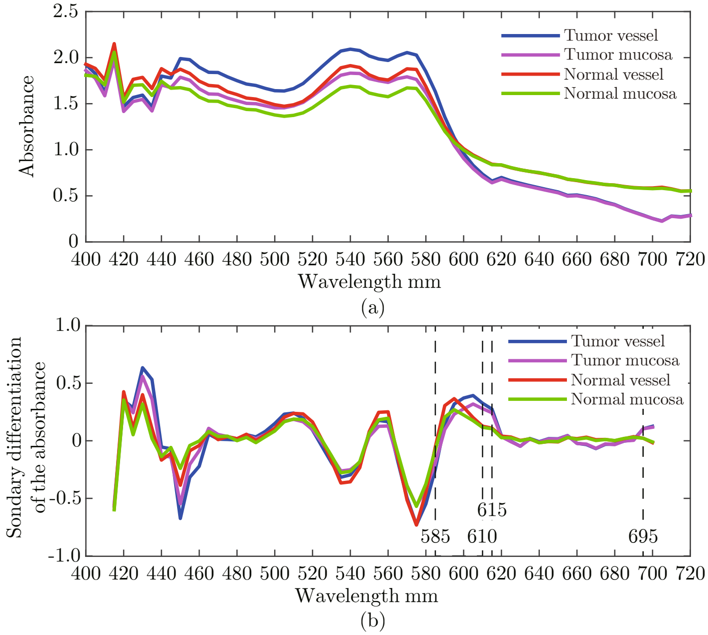


Fig. 4 The average absorbance and the average of the absorbance secondary differential value at each wavelength. Dash line shows the selected wavelength for extracting tumor

each the tumor vessel, the tumor mucosa, the normal vessel, and the normal mucosa using 540 nm narrow band image. In the 540 nm, that is the absorption peak of the hemoglobin, blood vessels are displayed black as compared with the mucous. So, the difference between the blood vessel and the mucous are emphasized. Then, we obtained the average of the absorbance secondary differential value by using Equation (2). Here, the value of $\Delta\lambda$ is 15nm. Then, we selected three wavelength using PCA for the tumor extraction. If the statistical significance was confirmed, the selected wavelength was used for correlation matching. Finally, we extracted tumor region by using Equation (8).

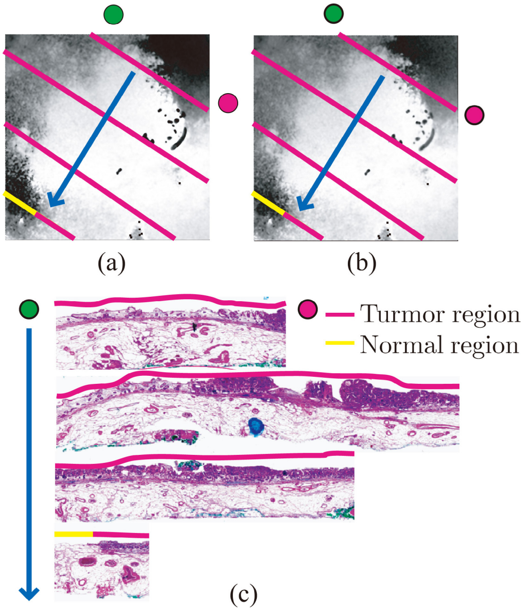


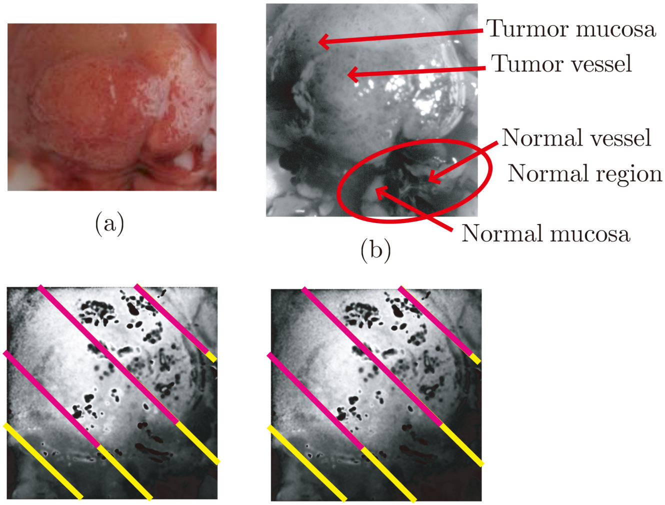
Fig. 5 Experimental results and the pathological results. (a) separation result by using three wavelengths (460nm, 550nm, 480nm), (b) separation result by using four wavelengths (adding 654nm), (c)pathological results. Pink lines show tumor region diagnosed by the pathological result and Yellow lines show normal region

3.2 Experimental Results

First target was the excised gastric polyp as shown in Fig. 3(a). Fig. 3(b) shows the narrow band image of this polyp at 540 nm. Based on this image, we selected 20 samples in each the tumor vessel, the tumor mucosa, the normal vessel, and the normal mucosa. The measured average absorbance and the average of the absorbance secondary differential value at each wavelengths show in Fig. 4 In order to examine the effect of the number of wavelengths, we compared the case of $L = 3$ and $L = 4$. In the case of $L = 3$, 460nm, 550nm and 480nm were selected. In the case of $L = 4$, 654nm was selected in addition to the above wavelengths. The extracting result shows in Fig. 5(a), (b). Fig. 5(a) shows the result of $L = 3$, and (b) shows one of $L = 4$. In this figure, the brightness represents the correlation strengths. The bright region corresponds to the tumor area.

Then we confirm the experimental result, tissues is examined by a pathologist. Fig. 5(c) shows the pathological specimen. Each cross-section corresponds to a diagonal cutting line of Fig. 5(a), (b). A pink line is in the tumor region and a yellow line in the normal region.

Second target was the excised colonic polyp as shown in Fig. 6(a). We conduct the experiment for this target in the same procedure. Fig. 6(b) shows the narrow band image of this colonic polyp. In the case of $L = 3$, 430nm, 445nm and 630nm



(a) (b)

(c) (d)

Fig. 6 Experimental results. (a) colonic polyp, (b) narrow band image at 540nm of (a), (c) separation result by using three wavelengths (430nm, 445nm, 630nm), (d) separation result by using four wavelengths (adding 700nm). Pink lines show tumor region diagnosed by the pathological result and Yellow lines show normal region

were selected. In the case of $L = 4$, 700nm was selected in addition to the above wavelengths. The extracting result shows in Fig. 6(c), (d). Fig. 6(c) shows the result of $L = 3$, and (d) shows one of $L = 4$. A pink line is in the tumor region diagnosed by the pathological examination and a yellow line in the normal region.

In Fig. 5(a), (b), compared the separation result by the proposed method and the result of the pathology, tumor region are generally separated. However, the tumor region of the upper right in the image are not separated. It is considered that the influence of the sprayed saline on the surface of the polyp. Meanwhile, the doctor assessed that this method separated tumor region of this sample with accuracy. The result in three wavelengths and one in four wavelengths is a similar, the difference in the results due to the difference in the number of wavelength was little. In Fig. 6(c), (d), compared the separation result by the proposed method and the result of the pathology, tumor region are also generally separated. However, there are some regions that are not able to separate. This is attributed to the effect of specular reflections and shadows. It suggest that it is necessary to devise the lighting, such as using polarized light. Meanwhile, according to the doctor, this method separated tumor region of this sample with accuracy. The result in three wavelengths and one in four wavelengths is a similar, the difference in the results due to the difference in the number of wavelength was little. The selected set of wavelength was different in two sample and so it is assumed that the optimal wavelength for separating is different in each sample. In the future, we need to verify it by increasing the number of samples. On the other hand, the number of wavelength did not affect the results.

From this result, it was confirmed that it is sufficient in the three wavelengths for separating tumor region. These results suggest that the region can be specified by using the proposed method.

4 Conclusion

In this paper, we have shown that the method for the separation of tumor region using multispectral image at optimum wavelengths automatically selected by PCA. In future, we would like to consider the optimization of obtained wavelengths for separating tumor region by increasing the number of samples.

References

1. Kiesslich, R., Fritsch, J., Holtmann, M., Koehler, H.H., Stolte, M., Kanzler, S., Nafe, B., Jung, M., Galle, P.R., Neurath, M.F.: Methylene blue-aided chromoendoscopy for the detection of intraepithelial neoplasia and colon cancer in ulcerative colitis. Gastroenterology **124**, 880–888 (2003)
2. de Veld, D.C.G., Skurichina, M., Witjes, M.J.H., Duin, R.P.W., Sterenborg, H.J.C.M., Roodenburg, J.L.N.: Clinical study for classification of benign, dysplastic, and malignant oral lesions using autofluorescence spectroscopy. J. Biomed. Opt. **9**, 940–950 (2004)
3. Mizuno, H., Gono, K., Takehana, S., Nonami, T., Nakamura, K.: Narrow band imaging technique. Tech. Gastrointest. Endosc. **5**, 78–81 (2003)
4. Gono, K., Obi, T., Yamaguchi, M., Ohyama, N., Machida, H., Sano, Y., Yoshida, S., Hamamoto, Y., Endo, T.: Appearance of enhanced tissue features in narrow-band endoscopic imaging. J. Biomed. Opt. **9**, 568–577 (2004)
5. Osawa, H., Yoshizawa, M., Yamamoto, H., Kita, H., Satoh, K., Ohnishi, H., Nakano, H., Wada, M., Arashiro, M., Tsukui, M., Ido, K., Sugano, K.: Optimal band imaging system can facilitate detection of changes in depressed-type early gastric cancer. Gastrointest. Endosc. **67**, 226–234 (2008)
6. Sigurdsson, S., Philipsen, P.A., Hansen, L.K., Larsen, J., Gniadecka, M., Christian Wulf, H.: Detection of skin cancer by classification of Raman spectra. IEEE Trans. Biomed. Eng. **51**, 1784–1793 (2004)
7. Paclk, P., Duin, R.P.W., Van Kempen, G.M.P., Kohlus, R.: Segmentation of multi-spectral images using the combined classifier approach. Image Vis. Comput. **21**, 473–482 (2003)
8. Iwahori, Y., Shinohara, T., Hattori, A., Woodham, R.J., Fukui, S., Behuyan, M.K., Kasugai, K.: Automatic polyp detection in endoscope images using hessian filter. In: Interantional Conf. Mach. Vis. Appl., pp. 21–24 (2013)
9. Leitner, R., Biasio, M., De Arnold, T., Dinh, C.V., Loog, M., Duin, R.P.W.: Multi-spectral video endoscopy system for the detection of cancerous tissue. Pattern Recognit. Lett. **34**, 85–93 (2013)
10. Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., Forman, D.: Global cancer statistics. CA. Cancer J. Clin. **61**, 61–69 (2011)

Analysis of Signal Variation Based on Path Loss in LTE Network

Zin Mar Myo and Myat Thida Mon

Abstract Long Term Evolution (LTE) network has a drastic development in the field of telecommunication broadband wireless network. It uses several propagation models which are available for various networks such as ad hoc networks, cellular networks, etc. Path loss is the significant concept to design and investigate the efficient signal propagation mechanism. In this paper, we give the analysis results about signal variation concerning with path loss over the LTE network. The analysis is based on an analytical model which is the use of processor sharing queuing theory. The analysis mainly considers at non-Guaranteed Bit Rate (non-GBR) such as FTP or HTTP. This paper also presents the work on the investigation of some parameter settings of the analytical model. Finally, the analysis of signal variation concerned with analytical model will be proved by comparing with the simulation result.

Keywords LTE · Path loss · Non-GBR

1 Introduction

Broadband wireless network is emerging drastically in the field of telecommunication in which Long Term Evolution (LTE) Technology is focused here. LTE is the wireless broadband technology to maintain handheld devices and cell phones to support internet access. LTE network provides high speed for data access for mobile phones. It may increase the speed and capacity of a network using different radio interface [1]. LTE technology has been developed with 3GPP

Z.M. Myo(✉)
University of Computer Studies, Yangon, Myanmar
zinmarmyomyo@gmail.com

M.T. Mon
University of Information Technologies, Yangon, Myanmar
myattmon@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing,*
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/ 978-3-319-23204-1_21

(3^{rd} Generation Partnership Project) to meet rapidly growing user requirements in telecommunication field [2] [3]. In LTE, there are two types of bearer [4]: Guranteed Bit Rate (GBR) which is real time traffic such as VoIP and non-Guaranteed Bit Rate (nonGBR) which is streaming such as data flow.

Fig.1 shows the general architecture of LTE network. The LTE system can be divided into two different evolutions mainly, the Enhanced Universal Terrestrial Radio Access Network (E-UTRAN) and the System Architecture Evolution (SAE). The LTE architecture can be said two nodes architecture. The two nodes are user equipment (UE) and evolved NodeB (eNB or eNodeB) which is base station.

LTE uses OFDMA in the downlink which means that the spectrum is divided into multiple subcarriers in the frequency domain and several OFDMA symbols in the time domain. The smallest unit defined within the LTE 3GPP specification that the scheduler can allocate over the radio is called Physical Resource Block (PRB). It consists of 12 subcarriers in the frequency domain and two slots in the time domain (i.e. 14 OFDMA symbols).

Fig. 2 shows the LTE downlink resource grid in over both time and frequency domains. Each subcarrier has 15 kHz bandwidth resulting in a PRB resolution of 180 KHz. This means that the LTE spectrum is divided into a number of PRBs. Table 1 shows the number of PRBs per each of the LTE transmission bandwidth. This number is not exactly the division of the spectrum by the 180 kHz since some of the subcarriers are reserved for signaling purposes.

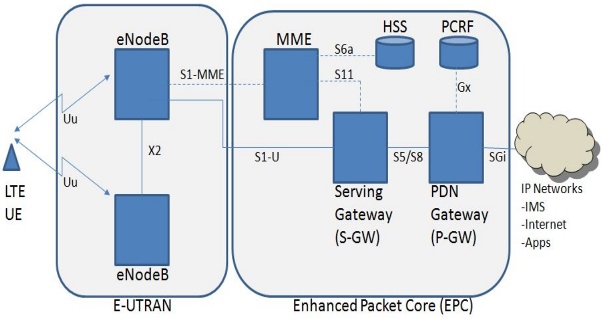


Fig. 1 LTE Network Architecture [5]

Path loss occurs when a signal propagates through air medium in NLoS. Path loss is calculated so as to analyze the link establishment in the telecommunication field. Link establishment can be calculated according to the gain and loss of the Transmitter (Tx) / Receiver (Rx) through air medium.

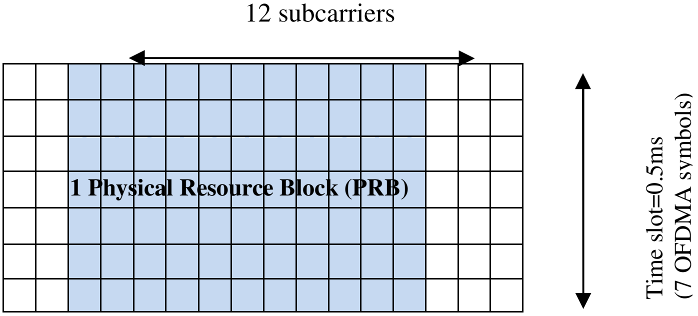


Fig. 2 LTE Resource Grid

Table 1 Number of PRBs per Different Spectrum

LTE Spectrum (MHz)	1.4	3	5	10	15	20
Number of PRBs	6	15	25	50	75	100

The paper is organized as follows: section 2 describes the works concerning with modeling and analysis on LTE radio scheduling and section 3 will explain about wireless channel model. Section 4 describes the considered the network scenarios and state the modelling assumptions. In section 5, analysis on LTE radion varion and investigation on some parameters are discussed and illustrated and finally section 6 will give the conclision and future work.

2 Related Works

Some of the previous researches works that are related to the proposed system are described in this section. Anton Riedl et al. [6] addressed the problem of dimensioning IP access networks for elastic traffic, which is carried by the TCP protocol. They employed M/G/R Processor Sharing model (M/G/R PS) as the basis for capacity computations and showed its applicability to TCP Internet traffic through simulations. In [7], Bartłomiej Błaszczyszyn and Mohamed Kadhem Karray showed how the downlink in LTE OFDMA cellular network can be dimensioned by means of a multi -rate Erlang's loss model that takes into account information theoretic foundations of the wireless communications.

Raymond Kwan et al. [8] discussed about allocating the resources to multiple users on the downlink of a Long Term Evolution (LTE) cellular communication system. Their numerical result showed that a limited amount of feedback information could provide a relatively good performance. The performance of a scheduler which addresses fairness among users was also presented. In [9], C. Wengerter et al. derived a closed-form expression for the throughput of proportional fair scheduling (PFS) in wireless networks. Then, an analytical model was given to provide the high accuracy of in evaluating the throughput of the PFS algorithm in Rayleigh fading environment. The theoretical results from analysis were compared with those from simulations. Dimitrova [10] presented the flow behavior analysis with the two scheduling schemes for LTE uplink. All users were scheduled in a Round Robin fashion and all were assigned an equal number of subcarriers regardless of the channel condition. The analysis was based markov chain. The author also investigated the impact of parameter settings of his analytical analysis.

3 Channel Model

Channel model is a mathematical representation of the relationship between power transmitted and received during a wireless transmission. Channel model describes the effect of channel on the transmitted signal. Fading in a channel is the variation of the strength of transmitted signal during propagation over time and over frequency. It shows the alteration of signal amplitude and phase caused by the transmission medium. Fading occurs due to several reasons including distance attenuation, obstacles, motion of transmitter and/or receiver and the arrival of multiple signals.

Fading encountered by a wireless signal is in general, an aggregate of three primary types of fading. These fading constituents are

- Path loss
- Slow fading
- Fast fading

Path loss and slow fading fall in the category of large scale fading. Large scale fading occurs if the mobile travels a distance which is comparable to the diameter of the cell. Path loss depends upon the distance between the transmitter and the receiver, while slow fading occurs typically due to shadowing effect when the signal is obstructed by large objects. Fast fading is categorized as small scale fading and occurs due to multipath propagation and motion of the mobile a distance comparable to the carrier wavelength [11].

4 Model and System Assumptions

There is only a single cell with N different zones of equal area in order to differentiate between user's distances to the base station. Mobile stations are assumed to be uniformly distributed over the cell zones and flow arrivals follow a Poisson process with rate λ. Hence the arrival rate per zone can be derived as $\lambda_i = {}^{\lambda}/_{N}$, where i= 1…N. Each zone is characterized by a distance d_i measured from the outer edge of the zone. We also assume that users don't move throughout the whole transmission.

Among three types of fading, the analysis mainly focuses on path loss. The signal deviation in each zone can be specified by the path loss $L(d_i)$. In the paper, the vehicular environment path loss model is chosen for analytical analysis. In [12], path loss for vehicular environment is formulated as

$$L(d_i) = 40(1 - 0.004 \times \Delta h_b)log_{10}(R) - 18log_{10}(\Delta h_b) + 21log_{10}(f_c) + 80 \quad (1)$$

Where, Δh_b is the height of the base station antenna in meters from the average roof top level, and $0 < \Delta h_b < 50$. For our analysis, setting the height of base station to 15m and the operating frequency to 2GHz gives the following path loss expression

$$L(d_i) = 37.6log_{10}(D) + 128.1 \quad (2)$$

Where, D is the distance between eNodeB and user equipment (UE). With the consideration of path loss, we define the received power at UE per zone as follows:

$$P_i^{rx} = \frac{P_{max}^{tx}}{M \cdot L(d_i)} \quad (3)$$

Where, p_{max}^{tx} is the maximum transmission power by eNodeB and $L(d_i)$ is the path loss and M is the number of PRBs distributed over the whole cell. Due to the nature of OFDMA, multipath interference can be assumed to be zero. As there is the single cell consideration, inter-cell interference cannot take into account. So only thermal noise N_0 is considered in the current model. The SINR (signal to noise and interference ratio) of each user located in the corresponding zones is

$$SINR_i = \frac{P_i^{rx}}{N_0} \quad (4)$$

Given the $SINR_i$ value, the data rate or channel capacity realized by a user when it is scheduled, which can be termed as instantaneous rate r_i, can be calculated according to the Shannon formula modified with a parameter σ [12].

$$r_i = (M_i(MS) * 180kHz)\sigma log_2(1 + SINR_i) \quad (5)$$

Where, r_i is the instantaneous data rate or the maximum throughput for the given $SINR$ and $M_i(MS)$ is the number of PRBs for all active users in a particular zone.

5 Analysis and Numerical Results

In the paper, FTP-only scenario as nonGBR is presented for the analysis. Since all of the flows (FTP downloads) can share the radio resources equally, a Processor Sharing (PS) model can be used to model the scheduler behavior. The propagation model is mainly based on path loss. Hence M/G/1-PS with path loss is a valid model for this situation. According to [7], the expected sojourn time (the average FTP downloading time τ) for a file with size χ in M/G/1-PS is given with the arrival rate λ :

$$E_{M/G/1-PS} = \frac{\chi}{C} \cdot \Delta 1 \rho \tag{6}$$

Where, ρ = traffic load, $\Delta 1$ = delay factor, and C = capacity value ρ is defined by the file size $\bar{\chi}$ and arrival rate λ as

$$\rho = \frac{\lambda \cdot \bar{\chi}}{C} \tag{7}$$

$$\Delta 1 \rho = 1 + \frac{Er_C(R, R \cdot \rho)}{R\,(1-\rho)} \quad |_{R=1} \tag{8}$$

$$= \frac{1}{1-\rho}$$

The mean traffic load and arrival rate in each zone are ρ_i and λ_i respectively. So the average file downloading time τ for a file size χ in zone i with $M/G/1-PS$ can be defined as

$$\tau = \frac{\chi}{r_i} \cdot \Delta 1 \rho \tag{9}$$

Where, r_i is capacity or instantaneous data rate in each zone.

5.1 Variation of Mean File Transfer Time (MFTT)

In this section, the variations of mean file transfer time due to path loss are discussed on the system of 10 MHz bandwidth. The distances of the different zones and number of PRBs for a particular zone are considered as shown in Table 2. The maximum transmission power by eNodeB is 20 watt (43dBm) which is appropriate for 10 MHz bandwidth. According to [13], the attenuation of implementation σ is taken as 0.6 for downlink scenario.

Fig. 3 shows the resulted changes of expected file downloading time with the consideration of rate $\lambda = 1$ and average file size $\bar{\chi}$ is 5Mbit. As equation 5 suggests, increasing the PRBs allocation leads to increase in realized data rate. Therefore, mean file transfer time (MFTT) of user close to the eNodeB is lower than those of remote one. The users close to cell edge have long time in download. Fig. 4 shows the comparison results of mean file delay time for the same consideration with no path loss consideration. The amount of throughput that is fallen due to path loss can be seen.

File download without path loss consideration is faster, but it is not realistic in any propagation model. In Fig 5, our analytical result is proved with simulation result by using LTE-Sim.

Table 2 Individual distance per zone and maximum PRBs allocation

Zone Number	1	2	3	4	5
Distance	0.3	0.5	0.65	0.85	1
$M_i(MS)$	50	40	30	20	10

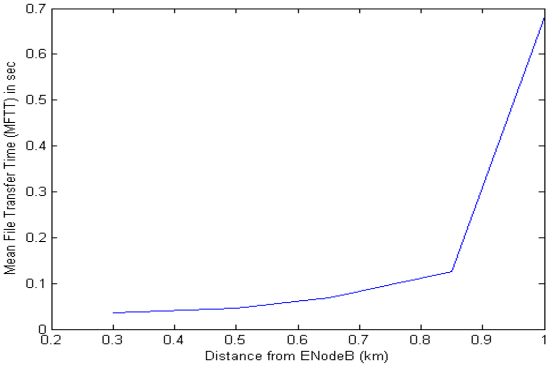


Fig. 3 MFTT Variation due to Distance Variation

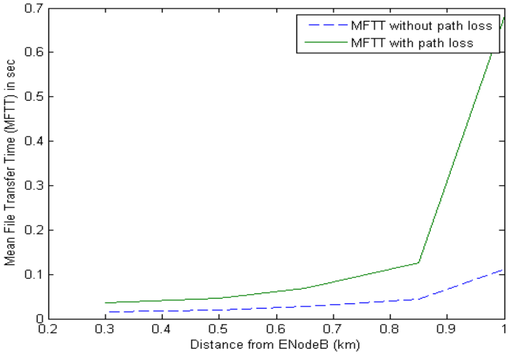


Fig. 4 MFTT Variation with and without path loss consideration

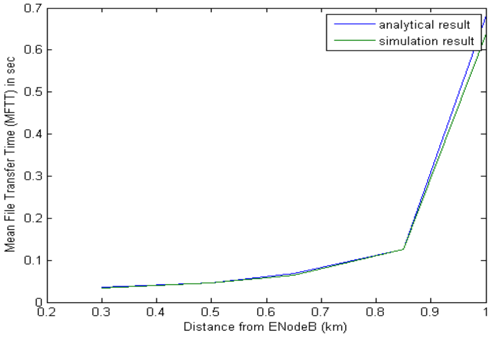


Fig. 5 Comparison between results from analytical model and simulator

5.2 *Impact of Parameterization*

Table 3 shows file transfer time between a range of arrival rates in 0.3, 0.5, 1, 1.5 flows /second and particular zone or assigned PRBs. As shown in the table, the value is negative at 1.5 flow/sec and 1km distance with 10 PRBs allocation.

Table 3 MFTT per different zone and different arrival rate

Distance	0.3 km (50 PRBs)	0.5km (40 PRBs)	0.65km (30 PRBs)	0.85km (20 PRBs)	1km (10 PRBs)
λ=0.3	0.0311	0.0401	0.0551	0.0870	0.2013
λ=0.5	0.0321	0.0418	0.0583	0.0953	0.2521
λ=1	0.0350	0.0467	0.0682	0.1250	0.6819
λ=1.5	0.0383	0.0528	0.0822	0.1819	**-0.9677**

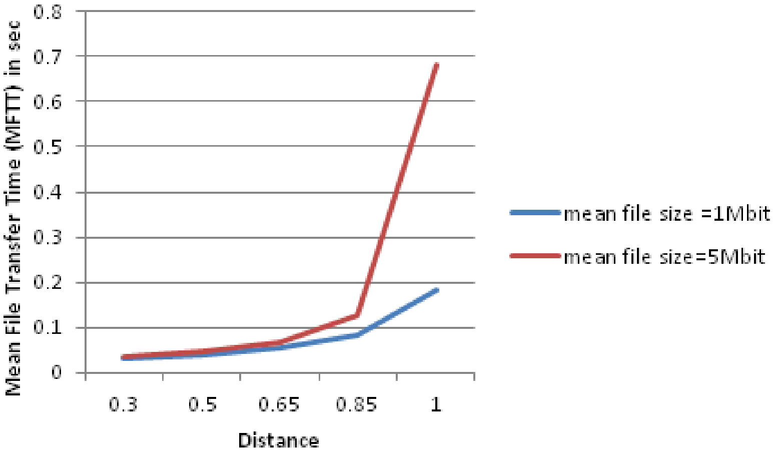


Fig. 6 Mean File Transfer Time for two different file sizes

This means that the traffic load is over flow in this area at this arrival rate. Therefore, the optimal choice of arrival rate should be dependent on PRBs allocation.

Fig. 6 shows the file transfer time for two different file sizes $\bar{\chi} = 1$ and $\bar{\chi} = 5$ at rate $\lambda = 1$. From this figure, the variation of file size has little impact on file transfer time at the distance that is not far from eNodeB. However, the significant difference in file transfer time can be seen at remote.

6 Conclusion and Future Work

In this paper, the analysis on signal variation due to path loss is made. This analysis is based on M/G/1-PS with path loss model. It considers at bit rate variation at nonGBR bearer on LTE scheduling downlink. Indeed, between path loss and fading which are the main features of radio variation, the model concerns just with path loss. Moreover, it can capture the condition of users are equally distributed in each zone. As an extended result, the paper discusses about the investigation on the system arrival rate and mean file size which are the parameters of traffic load. Analytical result is proved by comparing with simulation result also. Impact of PRBs allocation is going to be observed as a future work.

References

1. Ramkumar, J., Gunasekaran, R.: A New Path Loss Model for LTE Network to Address Propagation Delay. International Journal of Computer and Communication Engineering **2**(4), July 2013
2. Rao, M., Panchal, M.: 4G wireless technology: a survey paper. In: Proc. The National conference “NCNTE-2012”, pp. 44–47, February 2012
3. Shen, Z.K., Papasakellarious, A., Montojo, J., Gerstenberger, D., Xu, F.L.: Overview of 3GPP LTE advanced carrier aggregation for 4G LTE wireless communication. IEEE communication Magazine **50**(2), 120–130 (2012)
4. Alasti, M., Neekzad, B.: Clearwire Jie Hui, Rath Vannithamby, Intel Labs: Quality of Service in WiMAX and LTE Networks. IEEE Communications Magazine, May 2010
5. http://www.secforce.com/blog/2014/03/4g-lte-architecture-and-security-concerns/
6. Riedl, A., Baucschert, T., Perske, M., Probst, A.: Investgation of the M/G/R processor sharing model for dimensioning of IP access networks with elastic traffic. In: First Polish-German Teletraffic Symposium PGTS 2000,Dresden, September 2000
7. Błaszczyszyn, B., Karray, M.K.: Dimensioning of the downlink in OFDMA cellular via erlang’s loss model, In: Proc. of IEEE ITSC, USA (2009)
8. Kwan, R., Leung, C., Zhang, J.: Multi user scheduling on the downlink of and LTE cellular system, pp. 1–4 (2008)
9. Wengerter, C., Ohlhorst, J., von Elbwart, A.G.E.: Fairness and throughput an analysis for generalized propotional fair frequency scheduling in OFDMA. In: Vehicular Technology Conference (2005)

10. Dimitrova, D.C., van den Berg, H., Litjens, R., Heijenk, G.: Scheduling strategies for LTE uplink with flow behaviour analysis. In: 4th ERCIM workshop on eMobility, pp.15–26, May 31, 2010
11. Marwat, S.N.K.: LTE Channel Modeling for System Level Simulations. master mini project, September 2011
12. 3GPP technical Report TR 36.942. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios, Release 9 (2010)

Automatic Response Analyzer with Multiple Choices in Classroom Using Image Processing and Cards

Hiroshi Kamada, Kazuaki Masuda and Keisuke Yamada

Abstract To accelerate communication between a teacher and many students in classroom, we have proposed the system that can count students' raising the color cards processing images input with a web camera. Moreover, to improve recognition accuracy of the system, we have proposed the improved image processing system using the new color cards and multiple web cameras. In these systems, three color cards are used, and three answers can be chosen. In this paper, we propose the two systems with more choices than three. In the first system, four or five color cards would be used for four or five answer selection. In the second system, an asymmetry color pattern in a card would be used, and the rotating angle of a card would be used as an indicator of the answer. We used rectangular triangle color pattern in a card as the asymmetry color pattern in a card. We also report card recognition experimental results in the classroom.

1 Introduction

A teacher usually lectures many students in a classroom. To accelerate communication between a teacher and many students in a classroom, the conventional automatic response analyzers using the electronic wired or wireless terminals have been developed as the commercial products [1-4]. However, they are not popular, because they request large initial cost, labored preparation and maintenance. Special compact terminals as used in [1-2] must be managed carefully not to be lost or stolen. In case of commercial product [3], as all students may not have cell phones or smartphones, we must prepare the lacking cell phones or the smartphones. Students' PC terminals as used in [4] are not compact, and require IT classroom

---

H. Kamada(✉) · K. Masuda · K. Yamada
Graduate School of Engineering, Kanazawa Institute of Technology, Nonoichi, Ishikawa, Japan
e-mail: kamada@neptune.kanazawa-it.ac.jp

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/ 978-3-319-23204-1_22

with PC terminals. Furthermore, the conventional automatic response analyzers may result in sparse communication between a teacher and students, because the process does not require face-to-face communications. A manual method [5] was proposed that teachers count manually the color cards which students raised up, but it is time consuming to count the raising color cards. Based on the method [5], we have proposed the system [6-9] that can count automatically the students' raising color cards by processing the images input with the web camera. We show the concept of the system in Fig. 1. To improve recognition accuracy of the system [6], we have proposed the previous improved image processing system [7-9] using the new color cards and multiple web cameras.

2 The Previous System and Its Problem

The process of the system consists of image acquisition and image processing. We show the image input device [8-9] in Fig. 2. We have mounted two web cameras on the tripod, developing the connecting clasp. Since the two web cameras on the tripod are adjacent and higher than before, there is no un-captured area in front of area between two web cameras, and rear students' cards can be captured by the web cameras.

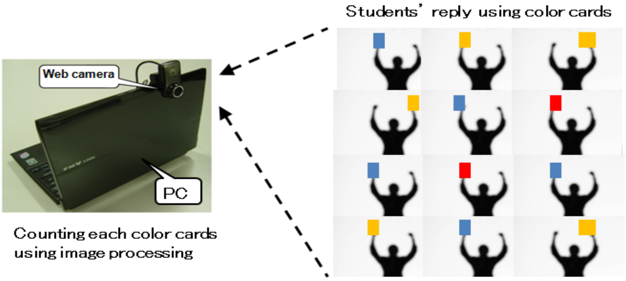


Fig. 1 The Concept of the System

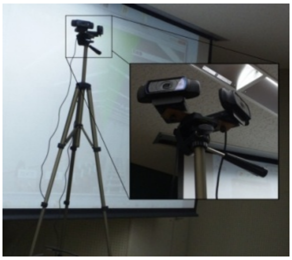


Fig. 2 The Image Input Device

We show A5-size cards [9] in Fig. 3. The color cards are made of two irreflexive papers, fluorescent colored paper and black paper. We show extraction process of four vertexes of colored area in a card at Fig. 4. At contour of binarized input image, vertex 1 is extracted as most far point from the searched first point at Cartesian coordinate, and vertex 2 is extracted as most far point from vertex1. Vertex 3 and vertex 4 are extracted as most far points from the segment connecting the vertex 1 and vertex 2. Matching between the square connecting vertexes 1, 2, 3, 4 and the contour of colored area is verified.

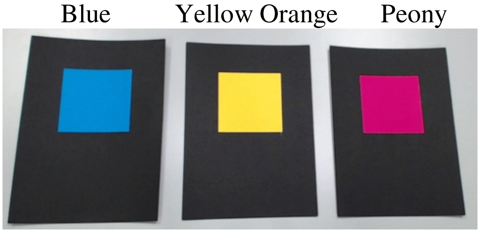


Fig. 3 The Color Cards

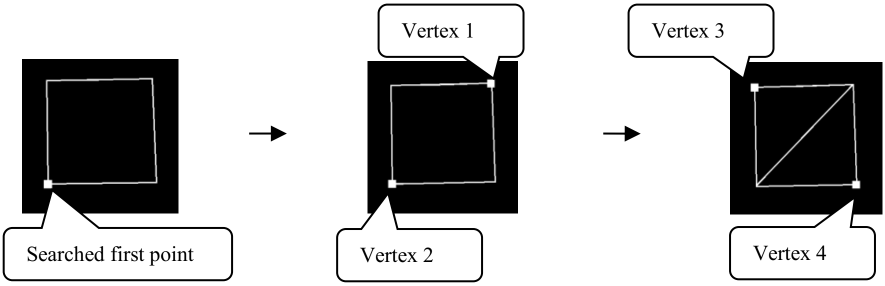


Fig. 4 Extraction Process of Four Vertexes of Colored Area in a Card

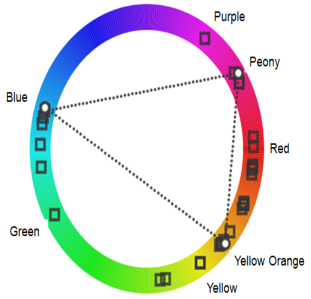
In these systems, three color cards are used, and three answers can be chosen. In the conversation, ‘Yes’ or ‘No’ are asked in most frequent pattern of questions. ‘A’, ‘B’ or ‘C’ is asked in second frequent pattern of questions. However, if the number of alternatives is two or three in the achievement test question, the correct answer may be selected with high probability 50% or 33% by chance, without understanding questions and answers.

3 The Systems with More Choices than Three

We propose the two systems with more choices than three. In the first system, four or five color cards would be used for four or five answer selection. In the second system, an asymmetry color pattern in a card would be used, and the rotating angle of a card would be used as an indicator of the answer.

3.1 The System Using Four or Five Kind of Color Cards

At first, we collected the 71 color papers, including papers with fluorescence paper of which the color is rather independent of lighting color. We didn't pick up printed color papers because the papers often reflect lighting. To discriminate color hues, the color saturation and lightness values must be high. Therefore, we selected 31 color papers with both of color lightness and saturation values no less than 150. We show the 31 candidate papers and their color values, hue, saturation and lightness measured using colorimeters in Appendix. Among the 31 colors, we selected seven candidate colors of which their hue values are distant each other, as shown in Fig. 5. Three colors jointed by dotted lines are used in the previous system.



Color name	Hue (0-179)	Saturation (0-255)	Lightness (0-255)
Yellow Orange	23	255	255
Yellow	30	255	255
Green	75	150	163
Blue	99	255	209
Purple	152	206	155
Peony	163	214	204
Red	0	160	255

Fig. 5 The Candidate Colors and Their Colors Values

3.2 The System Using Cards with Asymmetry Color Patterns

We selected the rectangular triangle color pattern in a card as the asymmetry color pattern in a card. We show the new card in Fig. 6. The new card is made as A4-size card, to make the color area be equal to previous A5-size card color area. Two rotation angle 0-degree and 180-degree may be represented as two answers. Their answers cannot be discriminated from all black backside of cards. We show extraction process of three vertexes of colored area in the new card at Fig. 7. At contour of binarized input image, vertex 1 is extracted as most far point from the searched first point at Cartesian coordinate, and vertex 2 is extracted as most far point from vertex1. Vertex 3 is extracted as most far points from the segment connecting the vertex 1 and 2. Matching between the triangle connecting vertexes 1, 2, 3 and the contour is verified. We selected the rectangular triangle color pattern in a card, as the two vertexes extraction processes shown in Fig. 4 and Fig. 7 are similar.

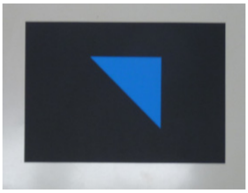


Fig. 6 The New Card with the Rectangular Triangle Color Pattern

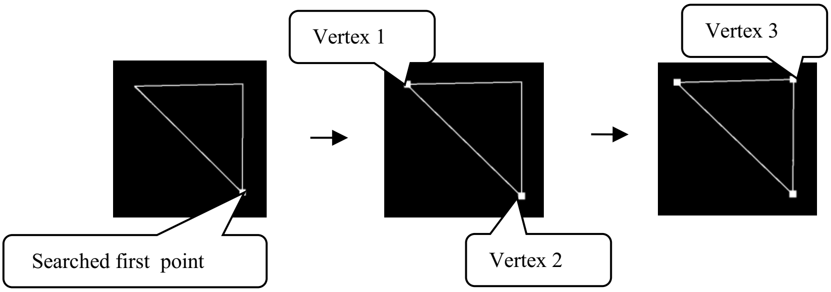


Fig. 7 Three Vertexes Extraction of Colored Area in the New Card

4 Evaluation Experiment

We have experimented in the classroom using the systems with more choices than three, described in chapter 3. We show the classroom specification in Table 1, and we show the classroom appearance and the seats layout in Fig. 8. The seating capacity of the classroom is 240 students. The floor of the classroom is terraced.

Table 1 The Classroom Specification

Width (m)	Depth (m)	Number of desks in width direction	Number of desks in depth direction	Number of seats in a desk
14.5	20.7	5	16	3

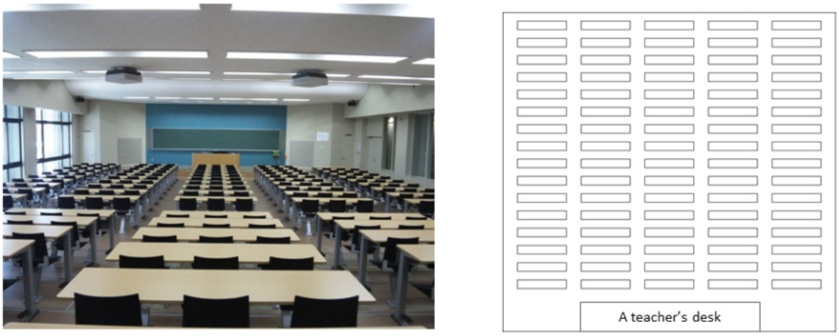


Fig. 8 The Classroom Appearance and the Seats Layout

We have put the 75 color cards on the desks in the classroom, one card per one desk except the front row desks, and we have experimented to recognize the color cards using the system. During the experiment we put out ceiling lights at the front row, or at the front row and the second rows, same as the situation when we use liquid-crystal projector to show the students PC screens. An example of the experimental result images in the classroom is shown in Fig. 9, putting out ceiling lights at the front row and the second row.

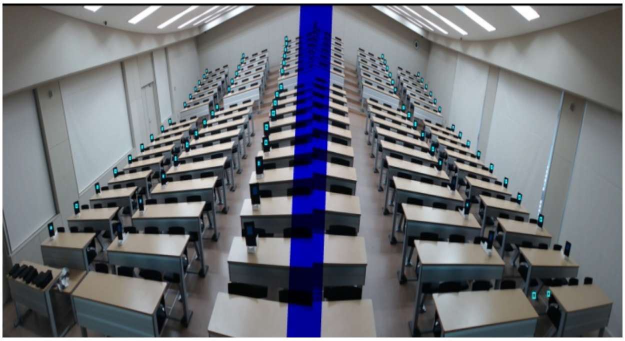


Fig. 9 An Experimental Result Image in the Classroom

Based on these candidate colors values in Fig. 5, we set the color values ranges shown in Table 2 to identify the candidate colors. Two ranges of saturation and lightness correspond respectively. We show the candidate colors cards recognition results in the Table 3. And, we selected representative three colors, and we have experimented about cards recognition using cards with the triangle patterns using the three colors as shown in the Table 4. The average processing time from image input to card recognition is 219 milliseconds using PC with 2.40GHz CPU.

Table 2 The Color Values Ranges to Identify the Candidate Colors

Color name	Hue (0-179)	Saturation (0-255)	Lightness (0-255)
Yellow Orange	10-30	128-, 50-	75-, 150-
Yellow	20-40	128-, 50-	75-, 150-
Green	50-90	128-, 50-	75-, 150-
Blue	95-115	128-, 100-	75-, 150-
Purple	125-160	128-, 100-	75-, 150-
Peony	155-172	128-, 100-	75-, 150-
Red	173-9	128-, 100-	75-, 150-

Table 3 The Candidate Colors Cards Recognition Results

Color name	Putting out ceiling lights at front row	Putting out ceiling lights at front row and second rows
Yellow Orange	100% (75/75)	99% (74/75)
Yellow	99% (74/75)	96% (72/75)
Green	59% (44/75)	47% (35/75)
Blue	100% (75/75)	89% (67/75)
Purple	92% (69/75)	68% (51/75)
Peony	100% (75/75)	72% (54/75)
Red	100% (75/75)	99% (74/75)

Table 4 Recognition Results of Cards with the Triangle Color Patterns

Color name	Putting out ceiling lights at front row	Putting out ceiling lights at front row and second rows
Yellow Orange	100% (75/75)	100% (75/75)
Blue	99% (74/75)	84% (63/75)
Red	97% (73/75)	97% (73/75)

5 Consideration

First, we consider about the system using four or five kinds of color cards. If colors values ranges to identify the colors overlap in Table 2, we cannot use the colors simultaneously. We show best cards colors selection for multiple choices in Table 5. Recognition rates about yellow orange and red cards are high. However, recognition rates about blue and peony cards are not so high. Recognition rate about green cards is low. Therefore, the system using four kinds of color cards is useful for appropriate use or lighting conditions. But it is difficult to construct the system using five kinds of color cards.

Table 5 Best Colors Selection for Multiple Choices

The number of kind of cards	Putting out ceiling lights at front row	Putting out ceiling lights at front row and second rows	Both two cases
3	Yellow Orange Blue Peony or Red	Yellow Orange Blue Red	Yellow Orange Blue Red
4	Yellow Orange Blue Peony Red	Yellow Orange Blue Peony Red	Yellow Orange Blue Peony Red
5	Yellow Orange Blue Peony Red Green	Yellow Orange Blue Peony Red Green	Yellow Orange Blue Peony Red Green

Next, we consider the system using cards with the triangle color patterns. As shown in Table 4, recognition rates about yellow orange and red cards are high. However, recognition rate about blue cards is not so high. Therefore, the system with four choices is useful using yellow orange and red cards. The system with six choices is useful for appropriate use or lighting conditions using yellow orange, blue and red cards.

6 Conclusions

We have proposed the two systems with more choices than three, and we have reported card recognition experimental results in the classroom. With respect to the system using four or five kinds of color cards, the system using four kinds of color cards is useful for appropriate use or lighting conditions. With respect to the system using cards with the triangle color patterns, the system with four choices is useful using yellow orange and red cards. The system with six choices is useful for appropriate use or lighting conditions using yellow orange, blue and red cards. Actual practical classroom application using the systems with more choices than three is the future study.

Acknowledgment This work was supported by JSPS KAKENHI Grant Number 24650571.

References

1. TERADA.LENON Co. Ltd.: LENON System. http://www.t-lenon.com/whatlenon.html
2. IC Brains Co. Ltd.: Socrates System. http://www.icbrains.com/soctop.html
3. CSE Co. Ltd.: mm-card System. http://www.cseltd.co.jp/mmcard
4. Computer Wing Co. Ltd.: Wingnet System. http://www.cwg.co.jp
5. Suetake, K.: A improvement of education in university using education engineering. Kanagawa University Engineering Laboratory Report (1989)
6. Kamada, H.: An Application of PC with Image Processing Accelerating Efficient Operation and Communication in Classroom. ICIC Express Letters Part B: Application (An International Journal of Research and Surveys) **6**(5), 1175–1180 (2012)
7. Kamada, H., Masuda, K.: An Automatic Response Analyzer in Classroom using Cards and Image Processing. ICIC Express Letters (An International Journal of Research and Surveys) **7**(5), 1681–1686 (2013)
8. Kamada, H., Masuda, K.: A Feasibility Study of Automatic Response Analyzer in Classroom Using Image Processing and Cards. ICIC Express Letters, Part B: Applications (An International Journal of Research and Surveys) **6**(4), 919–926 (2015)
9. Kamada, H.: R&D of simple interactive system and applying the system in the classrooms. In: Proceeding of Education Improvement Research Conference, pp. 90–91 (2014)

Appendix. The Candidate Papers and Their Colors Values

Color name	Hue (0-179)	Saturtion (0-255)	Lightness (0-255)
Reddish Orange[11]	3.5	178	255
Orange[1]	14	206	255
Yellow Orange[1]	**23**	**255**	**255**
Yellow[1]	**30**	**255**	**255**
Green[1]	40	255	255
Blue[1]	**99**	**255**	**209**
Purple[1]	**152**	**206**	**155**
Peony[1]	**163**	**214**	**204**
Peach[1]	165	188	255
Red[1]	178	175	255
Tomato[22]	4	188	255
Orange[2]	12	196	255
Mango[2]	24	229	255
Yellow[2]	30	255	255
Green[2]	42	216	255
Pink[2]	165	186	255
Red[2]	**0**	**160**	**255**
Vermillion[33]	4.5	168	239
Daidai[3]	5.5	153	255
Ocher[3]	11	153	221
Bright Yellow[3]	13	191	255
Orange[3]	13	193	255
Mandarin Orange[3]	20	170	255
Sunflower[3]	22	193	255
Lemon[3]	23	170	255
Green[3]	**75**	**150**	**163**
Emerald[3]	86	255	153
Pale Yellow[3]	91	255	153
Pale Blue[3]	96	255	209
Blue[3]	97	255	186
Red[3]	176	153	191

Colors with bold character are selected as the seven candidate colors.

---

[1] Illumi-colors #80 (fluorescent colored papers), Ehime Shiko Co. Ltd., Japan

[2] Illumi IJ #80 (fluorescent colored papers), Ehime Shiko Co. Ltd., Japan

[3] (No fluorescent colored papers), Daio Paper Corporation, Japan

Design and Implementation of Real Time Vision System for Recognizing Features of Lane Boundary

Aung Kyaw Tun, Kyaw Zaw Ye and L.G. Gagarina

Abstract A hard real-time vision system has been developed that analyses color videos taken from a car driving on a highway. This system uses a combination of color, edge, and motion information to recognize and track the road boundaries, lane markings and other vehicles on the road. A real time vision system detecting lane boundary and tracking vehicles on highway under reduced visibility is described and evaluated in this paper. By using a forward looking camera in a moving car the system recognize lane boundary, and the car driver is warned to the distance from other vehicle on the road and making decision of left or right lane departure. To decide controlling brake and turning left or right departure, the system uses the rear light signal of other vehicle on the road. In this research, color space transformation, edge detection, morphological boundary detection and spectrum difference are mainly used.

Keywords Image segmentation · Color space and transformation · Morphological image processing

---

A.K. Tun · K.Z. Ye(✉) · L.G. Gagarina
Department of Computer Science, Defence Services Academy, Pyin Oo Lwin,
Republic of Union of Myanmar
e-mail: Kyaw.Z.Ye@ieee.org

A.K. Tun · K.Z. Ye · L.G. Gagarina
Department of Informatics and Computer Software System,
National Research University of Electronic Technology, Zelenograd, Moscow, Russia
e-mail: kyawzawye85@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_23

1 Introduction

Vision systems that automatically analyze highway scenes from a moving vehicle have been remarkably being successful in fair weather conditions. Changes in road appearance due to weather conditions have been addressed for a stationary vision system that detects and tracks vehicles, for example, in reference [1,4, 5]. Reference [2, 6] describes how visibility leads to be reliable lane tracking from a moving vehicle. It recognizes and tracks lane marking on the road and multiple cars automatically without initialization by a human operator. The video data is processed in real time without any specialized hardware. All we need is an ordinary video camera and a low cost PC with an image capture board. The safety vehicle warning system can be used mainly in reduced visibility conditions such as rainy or foggy weather, tunnels and underpasses and driving at night dusk and dawn.

1.1 Motivation

Nowadays, computer technology is more widely used in vision based vehicle warning system for safety. Many dangerous conditions become rise in highway because of driving in excess speed, sleeping driver and poor visibility, etc. So, safety vehicle warning system gives aid to driver and passengers, and in developing of automatic driving.

1.2 Objectives of the System

The objectives of this research are as follows:

(a) To detect the lane boundaries of the road by Sobel edge detection method.
(b) To define features of lane edge by using morphological boundary detection methods.
(c) To warn about decision for overtaking and lane changing by using linear regression method.
(d) To evaluate the safety distance and lane departure from other vehicles by using color space differences.

2 Proposed System Architecture

The proposed system is as shown in figure.1. There are two main processes in my proposed system. They are morphological segmentation and color segmentation.

In my proposed system, input video is differentiated into morphological segmentation and color segmentation. In morphological segmentation, detect the image by changing from RGB to Gray method. Suppose the RGB value of a color is (r, g, b), where r, g, and b are integers between 0 and 255 [3]. The gray scale weighted average, x, is given by the formula

$$x = 0.299r + 0.587g + 0.114b. \quad (1)$$

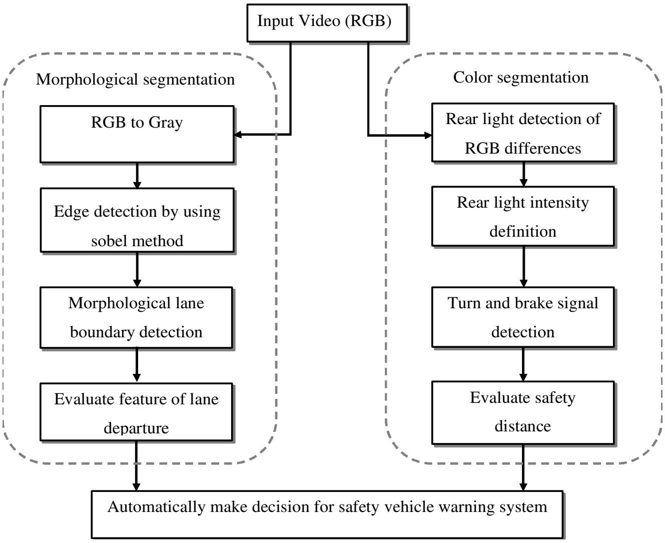

Input Video (RGB)
Morphological segmentation
RGB to Gray
Edge detection by using sobel method
Morphological lane boundary detection
Evaluate feature of lane departure
Color segmentation
Rear light detection of RGB differences
Rear light intensity definition
Turn and brake signal detection
Evaluate safety distance
Automatically make decision for safety vehicle warning system

Fig. 1 Proposed System Design

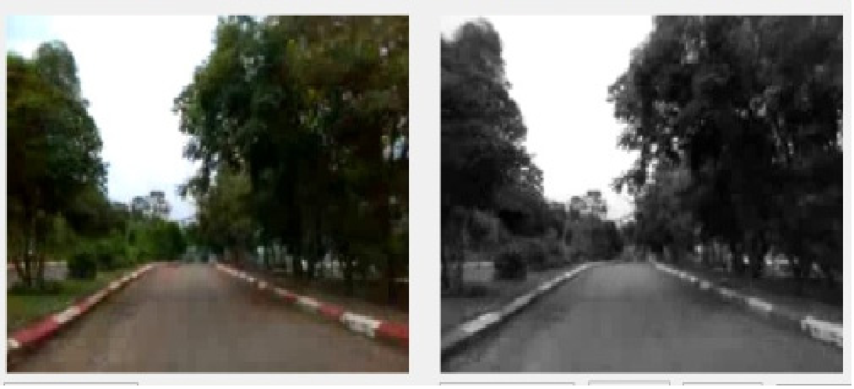

(a) (b)

Fig. 2 (a) RGB image (b) Convert to Gray image.

And then, with the aid of 'Sobel Operator' method, detect edge and make morphological detection. Sobel operator is a 3 x 3 neighborhood based gradient operator. The convolution masks for the Sobel operator are defined by the two kernels shown in figure 2. The two masks are separately applied on the input image to yield two gradient components Gx and Gy in the horizontal and vertical orientations respectively.

1	2	1
0	0	0
-1	-2	-1

-1	0	1
-2	0	2
-1	0	1

Fig. 3 Sobel masks

$$Gx=[f(i-1,j-1)+2f(i-1,j)+f(i-1,j+1)]-[f(i+1,j-1)+2f(i+1,j)+f(i+1,j+1)]$$

and

$$Gy=[f(i-1,j-1)+2f(i,j-1)+f(i+1,j-1)]-[f(i-1,j+1)+2f(i,j+1)+f(i+1,j+1)] \tag{2}$$

The gradient magnitude is usually computed as

$$G[f(x,y)]=\sqrt{Gx^2+Gy^2} \tag{3}$$

However, the other two formulae in equation. 2 can also be used to compute the gradient magnitude. The direction of gradient is computed as in equation.3.

The result of an edge image generated by the Sobel operator is shown in figure.4.

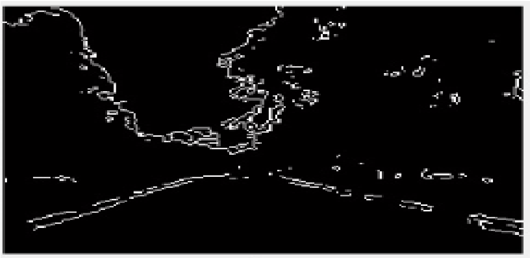


Fig. 4 Edge detects by using ‘Sobel Operator’.

Lane boundaries were detected by using morphological masks 4x4 matrix. There morphological masks 4x4 are as follow:

0	0	1	1
1	1	0	0
0	0	1	1
1	1	0	0

1	1	0	0
0	0	1	1
1	1	0	0
0	0	1	1

Mask-1 and Mask-2

0	0	0	0
0	0	1	1
1	1	0	0
0	0	0	0

0	0	0	0
1	1	0	0
0	0	1	1
0	0	0	0

Mask-3 and Mask-4

Fig. 5 Morphological Masks 4x4.

From that detects lane boundaries and finally evaluates features of lane departure.

The result lane boundaries image generated by the Morphological Masks is shown in figure 6.

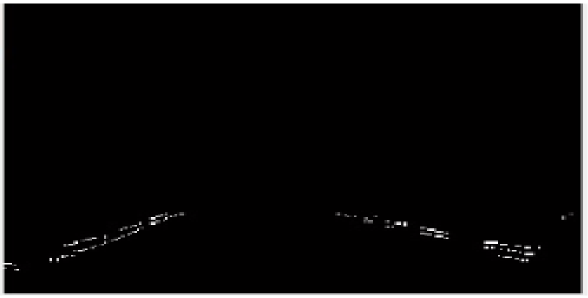


Fig. 6 Detection of lane boundaries.

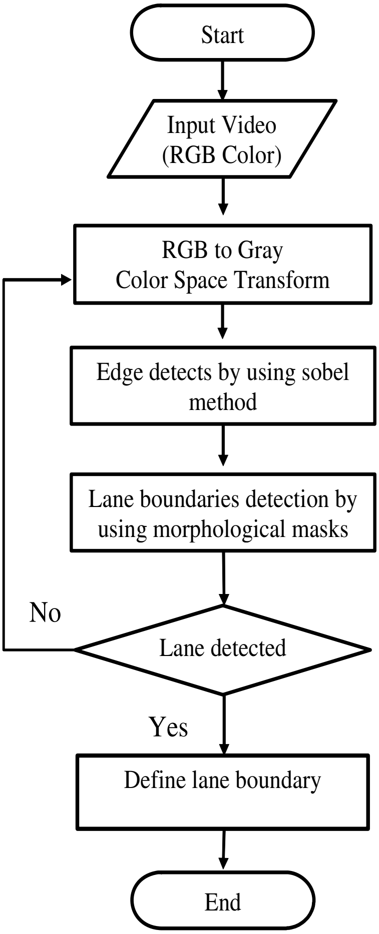

Start
Input Video (RGB Color)
RGB to Gray Color Space Transform
Edge detects by using sobel method
Lane boundaries detection by using morphological masks
No
Lane detected
Yes
Define lane boundary
End

Fig. 7 Flow Chart of morphological segmentation

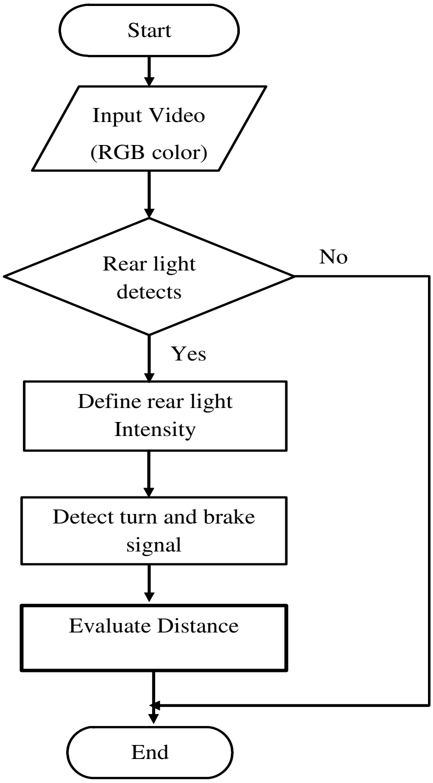

Start
Input Video
(RGB color)
Rear light
detects
No
Yes
Define rear light
Intensity
Detect turn and brake
signal
Evaluate Distance
End

Fig. 8 Flow Chart of color segmentation

Color segmentation were functioned as show in figure 7. In section of color segmentation, to evaluate safety distance, detect the rear light signals of other vehicles on the road. By using RGB difference, define the rear light intensity. Differentiate the turning and brake signal according to their color intensity and warn safety distance. By integration of evaluation lane boundary and safety distance, we can formulate and realize safety vehicle warning system.

3 Lane Boundary Detection

Now we have implemented lane boundary detection process as an application software system in MATLAB programming environment.

Firstly, input self-access video with resolution 180x240 pixels. Records the outline and number of frames and plays them. RGB images from every frame were converted into gray images. Edge detection was done by using 'Sobel Operator' method. After these steps detects the lane boundary by using assigned morphological masks 4x4.

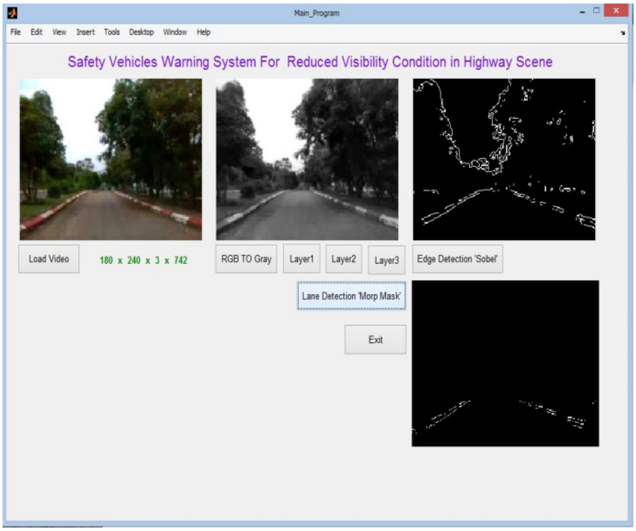

Main_Program
File Edit View Insert Tools Desktop Window Help
Safety Vehicles Warning System For Reduced Visibility Condition in Highway Scene
Load Video
180 x 240 x 3 x 742
RGB TO Gray
Layer1
Layer2
Layer3
Edge Detection 'Sobel'
Lane Detection 'Morp Mask'
Exit

Fig. 9 Lane Boundaries Detection System Interface

4 Conclusion

In this research work, we have presented the study of edge detection method, color space transformation and morphological image processing according to related background theory. We also presented the design of proposed system. We have used in this system 'Sobel Operator' method as edge detection and color space transformation from RGB TO Gray. We recognized features of lane boundaries by mask using morphological boundary detection method. In the future, we will evaluate features of lane departure and detect rear light, turn and brake signal by using RGB different method. We will construct rear light intensity definition upon our findings. By means of camera calibration method, we will evaluate safety distance and then finally make automatically decision.

Acknowledgements The research work was supported by National Research University of Electronic Technology and Laboratory of Information management systems and complexes, Moscow, Russian Federation.

References

1. Antonini, G., Sorci, M., Bierlaire, M., Thiran, J.: In: 8th International Conference on Advanced Concepts for Intelligent Vision Systems, 41, pp. 710–721. Lecture Notes in Computer Science. Springer Berlin (2006)
2. Bartlett, M.S., Littlewort, G.C., Frank, M.G., Lainscsek, C., Fasel, I.R., Movellan, J.R.: Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia **1**(6), 22–35 (2006)

3. Cheng, S.-C., Chen, M.-Y., Chang, H.-Y., Chou, T.-C.: Semantic-based facial expression ecognition using analytical hierarchy process. Expert Systems with Applications **33**(1), 86–95 (2007)
4. Acharya, T., Ray, A.K.: Principles and Application of image processing. Printed in the United States of America (2005)
5. Rojas, J.C., Crisman, J.D.: Vehicle detection in color images. In: Intelligent Transportation Systems. IEEE, Boston (1997)
6. Lin, F., Dong, X., Chen, B.M., Lum, K.-Y., Lee, T.H.: A robust real-time embedded vision system on an unmanned rotorcraft for ground target following. IEEE Transactions on Industrial Electronics, **59**(2), art. no. 5940220, 1038–1049 (2012)

Fish Classification Based on Robust Features Selection Using Machine Learning Techniques

Than Thida Hnin and Khin Thidar Lynn

Abstract The taxonomic identification of fishes is a time-consuming process and making errors is indispensable for those who are not specialists. This system proposes an automated species identification system to identify taxonomic characters of species based on specimens. It also provides statistical clues for assisting taxonomists to identify accurate species or review misdiagnosed species. For this system, feature selection is an essential step to effectively reduce data dimensionality. By using combination theory, this system creates the set of attribute pairs to construct the training dataset. And then each attribute pair in training dataset is tested by using two classifiers. Based on the accuracy result of each classifier on attribute pairs and the majority voting of each feature in these attribute pairs, this system selects the most relevant feature set. Finally, this system applied three supervised classifiers to verify the effectiveness of selected features subset.

Keywords Combination theory · Taxonomy · Identification · Fishes

1 Introduction

The pace of new species discovery and description would speed up significantly cause of multimedia and machine learning techniques could be developed. To automatically identify diagnostic features of specimens, simply choose between two alternatives at each step based on the presence or absence of a particular feature, the number of scales or the range of ratios between body measurements.

Recently, taxonomists have been searching for more efficient methods to meet species identification requirements, such as developing digital image processing and pattern recognition techniques. Researchers already used the pattern recognition techniques for insects, plants, spiders, and plankton. These approaches can be extended for fish. Although automated species identification might be a good option to

---

T.T. Hnin(✉) · K.T. Lynn
University of Computer Studies, Mandalay, Myanmar (Burma)
e-mail: {thanthidahnin,lynnthidar}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_24

the burden of routine fish taxonomic identification, there is not an automated taxonomic identification system for fishes based on specimen. In fact, automated species identification based on specimen has not become widely employed in any discipline of the biology. One of the explanations for why automated identifications have not become the norm for routine identifications is that such an approach is too difficult. The aim of this study is to determine whether morphometric variation among fish species allows automated taxonomic identification of the species. The key idea is to use the efficient machine learning algorithms for developing the fish identification system, rather than the ones used in traditional automated species identification systems. Machine learning algorithms are popular tools for classifying observations. These algorithms can attain high classification accuracy for datasets from a wide variety of applications and with complex behavior.

The success of applying machine learning methods to real-world problems depends on many factors. One such factor is the quality of available data. The more the collected data contain irrelevant or redundant information, or contain noisy and unreliable information, the more difficult for any machine learning algorithm to discover or obtain acceptable and practicable results. Feature subset selection is the process of identifying and removing as much of the irrelevant and redundant information as possible. Regardless of whether a learner attempts to select features itself, or ignores the issue, feature selection prior to learning has obvious merits.

Feature selection refers to the problem of selecting features that are relevant to predicting a target value for each instance in a dataset. Feature selection has several potential benefits: defying the curse of dimensionality to enhance the prediction performance, reducing measurement and storage requirements and reducing training and prediction times. This paper focuses on an attempt to maximize the performance of a classifier on unknown data and to recast the problem of feature selection in the context of taxonomic fish identification. To achieve efficient species identification, the proposed system can contribute in developing a system utilizing efficient features selection and classification techniques and provide automated fish identification system for Myanmar.

2 Related Work

Ecological interactions of fish assemblages in the pelagic environment can be partially determined by their larval distributions and recruitment to adult populations. The identification fish is essential for current studies on the distribution and reproductive strategies of pelagic fishes [2]. Thus, the assessment of biodiversity and its implication in the management of vulnerable marine ecosystems requires an accurate taxonomic identification of fishes. High levels of global biodiversity and a limited number of taxonomists represent significant challenges to the future of biological study and conservation. The main problem is that almost all taxonomic information exists in languages and formats not easily understood or shared without a high level of specialized knowledge and vocabularies. Thus, taxonomic knowledge is localized within limited geographical areas and among a limited number of taxonomists. This lack of accessibility of taxonomic knowledge to the general public has been termed the "taxonomic crisis" (Dayrat, 2005). Without this knowledge, the abundance of cryptic or unknown species might be under or overestimated. [1]

Meristic and morphometric characters are powerful taxonomic tools for measuring discreteness and relationships among fish species. For this reason, analysis of morphometric and meristic characters has not been widely used by ichthyologists to differentiate between different species and among different populations within a species. An automated species identification system is a matter of a one-to many matching, which not only needs to match an individual specimen with one of a set of extremely similar species to one another, but also is necessary to be able to reject it as belonging to a species that is not part of this set (Gaston and O'Neill, 2004) and patterns variation among fish species allows automated taxonomic identification of the species [5]. A family of automated species identification systems has been designed in recent years for gathering and analyzing data from images of specimens [7] [11].

3 Material and Methods

3.1 *Preprocessing*

Raw data is highly susceptible to noise, missing values and inconsistency. The quality of data affects the classification results. In order to improve the classification results raw data is preprocessed. Preprocessing is one of the most critical steps for machine learning algorithms as well as the data mining process which deals the preparation and transformation of the initial dataset.

In this proposed system, the data preprocessing has been performed in two steps. The data sets will used by this system are mixed of nominal and continuous types. However, most of the machine learning algorithms is primarily oriented to handling discrete features. Therefore, each continuous attribute is partitioned into categories. In the second phase, the features and the samples have been analyzed for missing variables and records with appropriate mean values by using mean imputation.

3.2 *Features Selection*

Most of the feature subset selection (FSS) algorithms are not suitable for all dataset. There is rarely a good way to choose appropriate FSS algorithms for various type of dataset. Therefore, FSS algorithm automatic recommendation is very important and practically useful.

Selecting suitable attributes is also an important step for effective and efficient fish classification. Many potential attributes may be used in the fish classification such as measurements and scale counts of body parts, and it can be done by the feature selection process. The main idea of feature selection is to determine the most relevant and the least amount of data representation of the specimen in order to minimize the within-class pattern variability, whilst, enhancing the between-class pattern variability.

In this paper, a combination based Feature Selection method is presented. It works as an iterative algorithm for feature selection, the Combination based

Feature Selection algorithm focus on optimizing the performance of the classifiers on unknown data. In this approach, feature vectors are first selected and then features are selected by using classification performance of classifiers and majority voting of the features as a criterion of feature subset selection.

3.2.1 The Proposed Combination Based Feature Selection Algorithm

The Combination based Feature Selection, described in detail in Fig. 1, is iterative in nature. It ranks each features vector according to its average accuracy value, and then selects the features vectors with the highest accuracy values

```
Find the optimal feature vector by evaluating the classification performance of classifiers for all possible feature vectors
Let D be the set of Fish data instances with m features, C be a set of predefined Classifiers
Let A be the feature set of each instance A={f1,f2,......fm}
Let k be the number of features and s be the feature vector with k features s={f1,f2,....fk}
Create S be the set of the feature vectors S = {s1,s2,....,sN} by using Combination Theory where N=$^{m}C_{k}$
// i.e m = 16, k = 4 , N=1840 length 4 vectors
Let S' be a set of selected feature vectors S'= ф
Optimal feature vector F= ф
For each classifiers c ϵ C
For each feature vector s ϵ S
        Measures the performance of each classifier c generated using D with s
    If classification accuracy greater than the accuracy-threshold
        S'= S' ∪ s          //select the relevant feature vectors
End for

For each feature fi in s in S'
    Compute voting(fj), the count of each feature in A that appeared in S', j=1,2,....m
voting(fj)=$\Sigma$ I(fj=fi)          //where I(.) is the indicator function that returns 1 if the argument is true and 0 otherwise.
End If
End for
Compute majority-voting (fi) =
    $= \frac{voting(fi)}{\sum_{i=1}^{m} voting(fi)}$
End For
Select the relevant features from A by using majority- voting of each feature fi that exceeds the voting-threshold
    F= F ∪{fi}
End For
Return F.
```

Fig. 1 Combination based Feature Selection Algorithm

3.2.2 Classification Algorithms

With the exponential growth in the amount of data that is being generated in recent years, there is a pressing need for applying machine learning algorithms to large data sets. Machine Learning algorithms are powerful tools not only for classification but also for the features selection. Most of these algorithms can get higher classification accuracy for datasets from a wide variety of bioinformatics applications with complex behavior as well as various application areas. This system considered two classification algorithms for evaluation of the relevant features subsets.

Naive Bayes Algorithm. Naïve Bayes is one of the most effective and efficient inductive learning algorithms. It has been widely used for data classification. As a classifier it learns from training data from the conditional probability of each attribute given the class label. Using Bayes rule to compute the probability of the classes given the particular instance of the attributes, prediction of the class is done by identifying the class with the highest posterior probability. Computation is made possible by making the assumption that all attributes are conditionally independent given the value of the class. Naïve Bayes as a standard classification method in machine learning stems partly because it is easy to program, its intuitive, it is fast to train and can easily deal with missing attributes [11]. For a Sample S with n levels of genus {g1, g2,….,gn} for the n features, the posterior probability that s belongs to class label c is

$$P(c|S) \sim \prod_{c \epsilon C} P(g_{1,}g_{2,} \dots, g_n|c)$$

Here $P(g_{1,}g_{2,} \dots, g_n|c)$are conditional probabilities estimated from training samples.

Attributes Selected Algorithm. This classifier can provide the automatically feature selection and classification procedure. It has two main functions (1) evaluation and (2) classification. The first step of this algorithm uses CFS to search feature subsets according to the degree of redundancy among the features. The aims to find the subsets of features that are individually highly correlated with the class but have low inter-correlation. To determine the best feature subset, this step is usually combined with best-first search strategies. The second step is the classification by using the result features subset.

4 The Proposed Automated Species Identification System

In this system, we use two classifiers to choose the main morphometric features of the fish species. These fish datasets are identified by the researcher of Mandalay University. Supervised machine learning classifiers such as Naïve Bayes (NB) and Attributed Selected (AS) that have been used many applications in bioinformatics are applied in feature selection. To maximize the performance of the classifiers on previously unseen data, and reducing training data, combination based feature selection is used. By using classifiers, the orders, families, genera and species can be discriminated based on the morphological characters (attributes).

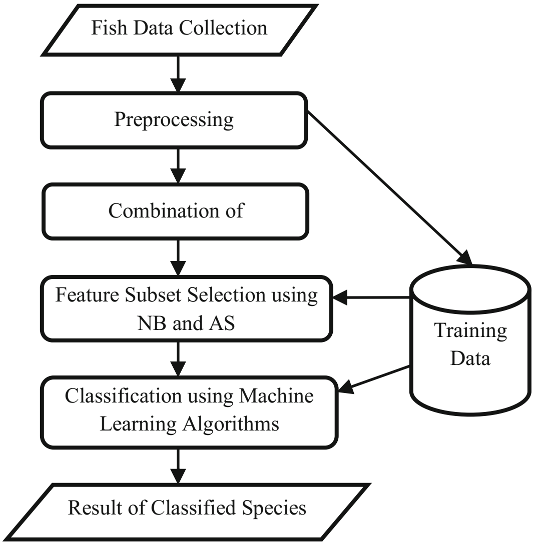

Fish Data Collection
Preprocessing
Combination of
Feature Subset Selection using NB and AS
Training Data
Classification using Machine Learning Algorithms
Result of Classified Species

Fig. 2 System Flow Diagram of the system

4.1 Dataset Description

Recent years, some researchers (taxonomists) from Mandalay University attempt to build the taxonomic fish identification system. However there are some inherent problems occur in the features selection and classification. The first important limitation in the dataset is the huge number of records. And researchers could not create the efficient database for these data sets. Researchers worked manually on documents and could not effectively identify because of the lack of methods and huge amount of data sets. The proposed system has to create fish database and use this database for determining the main taxonomic features of fish that are promoting divergence among closely related species.

4.2 Feature Evaluation

The parameters used to fine tune the feature selection are accuracy-threshold (at) and voting-threshold (vt).

Only the accuracy and frequency measures which exceed these predefined thresholds are considered for evaluation. On testing with various values for these parameters we found the system gave optimum performance for the values of at=0.9 and vt=0.043.

4.3 Experiment Results

This system performs the feature subset selection by using different classifiers and combination theory. In this work, we used the fish datasets of 1516 instances belonging to 20 classes. Each instance contains 16 attributes. These attributes are Mouth, Teeth, Barbels, Snout, Operculum, Eye, Head, Predorsal Scales, Dorsal Fin, Pelvic Fins, Pectoral Fins, Anal Fin, Caudal Fin, Dorsal Fin Spines, Adipose Fin and Lateral Line. Three popular classification algorithms including J48 decision tree-inducing algorithm, Multilayer Perceptron (MLP) and Support Vector Machine (SVM) are tested with different number of features on Fish dataset. For the purpose of this experimentation, Weka, Data Mining open source machine learning software is used [5]. Parameter settings were changed only when a significant improvement in performance based on preliminary experimentation was obtained. For the MLP learner, hidden layers were changed to 3 to define a network with one hidden layer containing three nodes. For SVM, the complexity constant c was changed from 1.0 to 3.0.

The evaluation of the best features subsets of this system required comparing performance over all possible subsets of features on each classifier. This estimation considered both large and small features sets to estimate the performance of classifiers for the task of feature selection. And we assumed that the size of significant interactions between different combinations of features is much smaller than 16 features, we limited ourselves to evaluating the performance from combination sampled from the whole set of features, with being a bound on the combination with k=4. The results along with the experimentation of different methods are compared based on accuracy, F-measure, area under the ROC, average precision and True Positive rate, False Positive rate and Recall. Table 1 summarizes the classifiers' performance on the test set and the number of features selected in each of the experiments. The accuracy levels are the fraction of correctly classified test set instances. The result shows that for most of the classifiers tested in this experiment, the features selected by proposed method work better in accuracy than the features selected by Correlation based Feature Selection method. Especially, the proposed method can effectively reduce the irrelevant features and successfully compete with an existing feature selection method. It can select the 5 relevant features, Mouth, Teeth, Snout, Head and Dorsal fin spines.

Table 1 Comparison of Accuracy levels and number of features selected using the different Classifiers

Classifiers	16 Features (without Features Selection)		(CFS with 9 Features)		(Proposed with 5 Features)	
	Accuracy	Time(Sec)	Accuracy	Time(Sec)	Accuracy	Time(Sec)
MLP(ANN)	79.1304%	10.64	79.1304%	6.84	82.6087 %	4.61
SMO(SVM)	100%	3.58	99.1304%	3.6	99.1304%	2.73
J48	92.1739%	0.07	92.1739%	0.05	93.0435%	0.03

5 Conclusion

In this paper, combination based features selection method is used. This method achieved not only the features subset with less features but also the result of classification accuracy. It can also be compared with the existing correlation based features selection method. The experimental results showed that the combination based feature selection methodology can define the curse of dimensionality to enhance the prediction performance, reduce measurement and storage requirements and also reduce training and prediction times. These results successfully demonstrated the value of applying combination theory concepts and majority voting method to feature selection. By using this feature selection method, we can extract the 5 common features. We used these best features for constructing the training examples with good generalization capability to correctly classify the unknown class label of instance.

The key idea of this system is to reduce the time spent on the taxonomic identification of fishes and to provide a tool for accurate classification. For future work, further classifications are required to observe the feature subset selection in gene expression and to identify the fish species accurately and automatically based on different machine learning techniques.

References

1. Hernández-Serna, A., Jiménez-Segura, L.F.: Automatic identification of species with neural networks. Hernández-Serna and Jimenéz-Segura. PeerJ (2014). doi:10.7717/peerj.563
2. Taha, A.M., Mustapha, A., Chen, S.-D.: Naive Bayes-Guided Bat Algorithm for Feature Selection. Hindawi Publishing Corporation. The Scientific World Journal, vol. 2013, Article ID 325973
3. Boulesteix, A.-L., Janitza, S., Kruppa, J., König, I.R.: Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics, July 25, 2012
4. Karegowda, A.G., Manjunath, A.S., Jayaram, M.A.: Comparative Study of Attribute Selection Using Grain Ratio and Correlation Based Feature Selection. International Journal of Information Technology and Knowledge Management **2**(2), 271–277 (2010)
5. Guisande, C., Manjarrés-Hernández, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuna, A., Prieto-Piraquive, E., Janeiro, E., Matías, J.M., Patti, C., Patti, B., Mazzola, S., Jiménez, S., Duque, V., Salmerón, F.: IPez: An expert system for the taxonomic identification of fishes based on machine learning techniques. Fisheries Research **102**, 240–247 (2010)
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The weka data mining software: an update. ACM SIGKDD Explorations Newsletter **11**(1), 10–18 (2009)
7. Chen, H., Bart Jr., H.L., Huang, S.: Integrated Feature Selection and Clustering for Taxonomic Problems within Fish Species Complexes. Journal of Multimedia **3**(3), July 2008

8. Inza, I., Calvo, B., Armañanzas, R., Bengoetxea, E., Larrañaga, P., Lozano, J.A.: Machine Learning: An Indispensable Tool in Bioinformatics
9. Ali, J., Khan, R., Ahmad, N., Maqsood, I.: Random Forests and Decision Trees. IJCSI International Journal of Computer Science Issues **9**(5), No 3, 1694–0814, September 2012
10. Breiman, L.: Random Forests. Machine Learning **45**(1), 5–32 (2001)
11. Alsmadi, M.K., Omar, K.B., Noah, S.A., Almarashdeh, I.: Fish Recognition Based On Robust Features Extraction From Color Texture Measurements Using Back-Propagation Classifier. Journal of Theoretical and Applied Information Technology (2010)
12. Abraham, R., Simha, J.B., Iyengar, S.S.: Effective Discretization and Hybrid feature selection using Naïve Bayesian classifier for Medical data mining. International Journal of Computational Intelligence Research **5**(2), 116–129 (2009). ISSN 0974-1259
13. Kohavi, R.: Scaling up the Accuracy of Naïve-Bayes Classifier: Decision Tree-Hybrid
14. Kamruzzaman, S.M., Haider, F., Hasan, A.R.: Text Classification using Association Rule with a Hybrid Concept of Naive Bayes Classifier and Genetic Algorithm
15. Rao, T., Rajinikanth, T.V.: A Hybrid Random Forest based Support Vector Machine Classification supplemented by boosting. Global Journal of Computer Science and Technology, C Software and Data Engineering **14**(1), Version 1.0 (2014)
16. Talwar, P.K., Jhingran, A.G.: Inland Fishes of India and adjacent Countries, vol. I, II. Oxford and IBH Publishing Co. Ltd., Calcutta, pp. 1–1158

Multi-face Tracking with Occlusion Recovery

Yea-Shuan Huang and Chin-Ie Chang

Abstract This paper proposes a face tracking method which uses the concepts of multiple instances and Online Adaboost to successively train and update a face tracking model (FTM) for each tracked face, and then tracks the tracked faces among the image frames. In order to avoid the continuous accumulation of tracking error, this method performs a face detection process around each tracked face. This paper also provides a novel tracking recovery approach when the tracked faces are occluded with other faces. First, it uses an occlusion detection criterion to realize whether a tracked face is occluded by other objects (called shelter) or not. If it is, several candidate regions around each shelter will be constructed, and by using frame difference and skin detection information, the candidate regions will be further filtered out so that the regions with no face clues will be omitted. For the remaining regions, it will apply the FTM preserved just before occlusion to detect whether a tracked face has reappeared or not. If an occluded face indeed has reappeared, it will continue to track this face from its reappeared position. Experiments show that the proposed method provides very good face tracking results on different videos and outperforms both Camshift and Kalman Filters.

1 Introduction

Face tracking is an essential technique to video surveillance and it can be used to lock each already detected face and locate their positions and extents in the successive frames. Till now, several object tracking methods have been proposed and can be mainly attributed into two categories. The first category is to individually locate the changing parts (called "foreground part") in images, and then through a feature matching operation, each tracked object find its corresponding one among all the found foreground parts in each successive image. The second category is to find the most possible location of each tracked object in a search region of the successive image by using the previously constructed object template. Some commonly-used methods of the first category are Background

---

Y.-S. Huang(✉) · C.-I. Chang
CSIE, Chung Hua University, Hsinchu, Taiwan
e-mail: yeashuan@chu.edu.tw

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_25

Subtraction [1], Temporal Differencing, and Optical Flow [2]. Among these methods, Background Subtraction is the simplest and most popular one. In order to accommodate background variation, a background model in general should be updated from time to time. The Codebook technique [3] is a commonly-used algorithm to construct an adaptive background model by using both quantization and clustering processes.

The second category relates to object tracking and matching, and it in general can be divided into two kinds of approaches. The first kind finds the corresponding position between two neighboring image frames, and the second one defines the regions and features of tracked objects, and with an either model-based or template-based method, it locates the corresponding image regions having the defined object characteristics. Article [4] belongs to the first-kind approach which finds the best matched region for each moving object by using a block matching algorithm, computes the motion vector of each matched region, and assigns the most frequently occurring motion vector to be the object motion vector. Mean Shift[5], Particle Filter [6][7] and Online-MilBoost [8] belong to the second-kind approach, and they all find the best tracking locations with an iterative manner. Some object tracking methods have attempted to include object detection recently and has obtained promising tracking result. Online MiBoost is one of such tracking algorithms. Article [9] shows Online-MilBoost has obtained good tracking performance.

This paper in total contains 5 sections. Section 2 Introduces the basic concept of the OnLine-MilBoost algorithm. Section 3 describes the proposed face tracking method which closely integrates face detection into face tracking operation. Section 4 compares the experimental results of both the proposed method and some other famous tracking methods. Finally, Section 5 draws a conclusion of this paper.

2 Object Tracking by Online MilBoost

The Online MilBoost algorithm adopts two important concepts, multiple instance learning (MIL) and online Adaboost. With MIL, useful features for object tracking are extracted from samples automatically. In order to obtain enough timely samples suitable for on-line training, a random sampling approach around the current tracked object is used. Suppose in one image, l_t^* is the center position of the tracked object, x is an image patch of a tracked object, y is a binary variable indicating the presence of the object of interest in image patch x, and $l(x)$ denotes the location of x. Two sets of image patches will be cropped from this frame, called positive bag and negative bag, as

$$\text{Positive bag}: \quad X^r = \{x \mid r > \| \, l(x) - l_t^* \, \| \} \tag{1}$$

$$\text{Negative bag}: X^{r,\beta} = \{x \mid \beta > \| \, l(x) - l_t^* \, \| > r\} \tag{2}$$

here r and β are two values to indicate the range of cropped samples and $r < \beta$.

MIL uses a randomized Harr feature set to train the tracked object model. There are a finite number of Harr feature groups, and each feature group consists of several Harr features with random parameter setting including feature number, position, size, and weight. Randomized Harr features can effectively preserve an image properties and they are obtained by the following manner.

$$f(x) = \Sigma_{i=1}^{n} w_i \cdot Rect_i(x) \tag{3}$$

$$n = Rand(p, q) \tag{4}$$

$$w_i = Rand(-1,1) \tag{5}$$

$$Rect_i = Rand(x, \text{image\_width}, \text{image\_height}) \tag{6}$$

here, n is an integer randomly selected between two given integers p and q; $f(x)$ denotes one feature group of image patch x which contains n features; w_i is a weight randomly assigned to the i-th feature of a feature group and its value is between -1 and 1. Then, an Online-MILBoost algorithm by associating both Multiple-Instance learning and Online-Adaboost learning is used to track human faces. This algorithm selects individual feature groups one after another from the randomized Harr feature set and gradually constructs a strong face detector. The Online-MILBoost algorithm is described as follows..

```
Algorithm Online - MILBoost
Input : Dataset$\{X_i, y_i\}_{i=1}^{N}$, where $X_i = \{x_{i1}, x_{i1}, \ldots x_{ij}\}_{j=1}^{P_i}$, $y_i \in \{0,1\}$
Train all M weak classifiers in the pool with data $\{x_{ij}, y_i\}$,
Calculate $\mu_0, \sigma_0, \mu_1, \sigma_1$
Initialize $H_{ij} = 0$ for all $i,j$
for k = 1 to K do
    for m = 1 to M do
        $p_{ij}^{m} = \sigma(H_{ij} + h_m(x_{ij}))$
        $p_i^{m} = 1 - \Pi_j (1 - p_{ij}^{m})$
        $\varepsilon^{m} = \sum_i (y_i \log(p_i^{m}) + (1 - y_i) \log(1 - p_i^{m}))$
    end for
    $m^* = \mathrm{argmin}_m \varepsilon^{m}$
    $h_k(x) \leftarrow h_{m^*}(x)$
    $H_{ij} = H_{ij} + h_k(x)$
end for
OutPut : Classifier $\mathbf{H}(x) = \sum_k h_k(x)$,    $p(y \mid x) = \sigma(H(x))$
where $\sigma(x) = \frac{1}{1 + e^{-x}}$
```

According to the location and size of a tracked face appeared in the previous frame, a search region is set for the current frame, and the possible face patches are obtained randomly within the search region. Then, the features adopted in the FTM are used to compute the feature values of each possible face patch, which are further inputted to the constructed strong classifier and the corresponding output represents the possibility score that the image patch contains the tracked face. Among the scores of all image patches, the one having the highest score is selected to be the tracked result if this score is larger than a predefined threshold.

In a real-time face tracking system, the tracked face model should be updated accordingly to accommodate different variations (such as changes in face orientations and illumination). After tracking a new face location, the FTM is updated as follows

$$\mu_1 \leftarrow (1-\gamma)\mu_1 + \gamma \cdot \frac{1}{n}\sum_{i|y_i=1} f_k(x_i) \tag{7}$$

$$\sigma_1 \leftarrow (1-\gamma)\sigma_1 + \gamma \cdot \sqrt{\frac{1}{n}\sum_{i|y_i=1}(f_k(x_i) - \mu_1)^2} \tag{8}$$

here, γ is an updating rate. The updates of μ_0 and σ_0 are the same as those of μ_1 and σ_1. After obtaining a new location of the tracked face, among the original M feature groups the K best ones are selected sequentially to form weak classifiers, and finally a new strong classifier is constructed for the subsequent frame tracking.

3 Method for Human Face Tracking

3.1 Single Face Tracking

Single face tracking does not mean only one face can appear in one frame, instead, it means a tracked face does not overlap with other objects in the current frame. In total, there are eight steps in the single face tracking process as shown in Fig. 1 and they are

Step 1. Input an image;

Step 2. Check whether the process is in a tracking mode, if it is, jump to step 5; otherwise, it means there is no tracking target, then jump to step 3;

Step 3. Apply face detection to the entire image and if there is no face detected, jump to step 1.

Step 4. Train the tracking face model, and jump to step 1;

Step 5. Perform face tracking, then perform face detection in the neighborhood of the tracked face. If there is no face detected, jump to step 7

Step 6. Calculate the distance between the locations of the tracked face and the detected face. When the distance is larger than a threshold, use the detected face information to update the tracking model of the tracked face and jump to step 8.

Step 7. Use the tracked face information to update the tracking model of the tracked face.

Step 8. Check whether the tracked face is disppeared or not, if it is, stop tracking this face and reset the model to be a non-tracking model, otherwise jump to step 1.

Step 5 performs a face detection around the location obtained by the FTM tracking so that possible face drifting can be detected and corrected during tracking. But with the help of face detection, they can be updated to resume proper tracking.

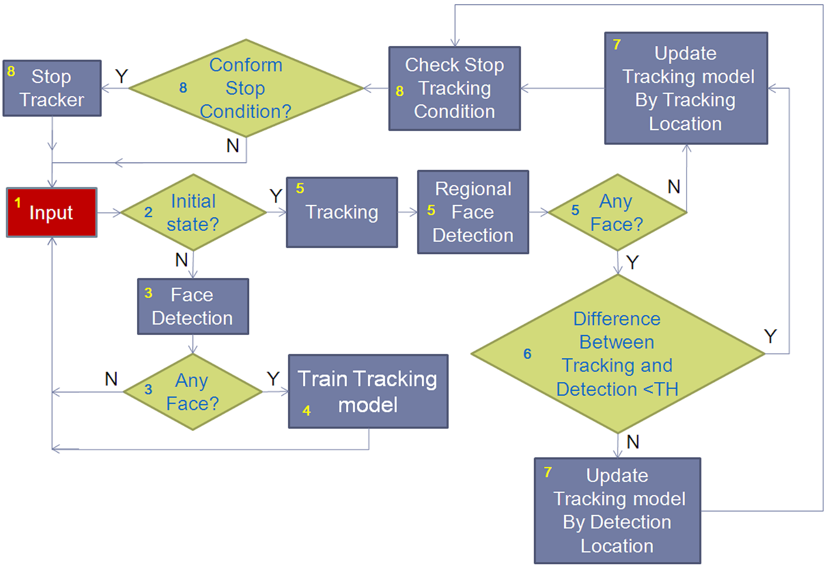

8 Stop Tracker
Y
8 Conform Stop Condition?
N
8 Check Stop Tracking Condition
7 Update Tracking model By Tracking Location
1 Input
2 Initial state?
Y
N
5 Tracking
5 Regional Face Detection
5 Any Face?
N
Y
3 Face Detection
N
3 Any Face?
Y
4 Train Tracking model
6 Difference Between Tracking and Detection <TH
Y
N
7 Update Tracking model By Detection Location

Fig. 1 Flow chart of single human face tracking

In a tracking system, it is necessary to have a mechanism to determine whether a tracked face has disappeared from the screen or been incorrectly tracked already. To this end, a stop conditions to terminate a face tracking are proposed. When the tracking score is smaller than a threshold value, the score counter will then be incremented by one, otherwise, the score counter will be reset to zero. Then, if the score counter is larger than a predefined score counter threshold (e.g. 10), tracking will be stopped.

3.2 Multiple Face Tracking

This section will introduce how to track faces when more than one face appears in one image frame, and how to resume tracking of faces when they are occluded by other objects and reappear again.

3.3 Process Flow of Multiple Face Tracking

Fig. 2 shows the process flow of the proposed multiple-face tracking system. If a tracked face is not occluded by other objects, it can be treated with the mentioned single face tracking algorithm. In order to realize a new person has entered the scene, a global face detection operation will be performed on the entire image for every several frames. Then in the same frame, all detected face regions will be checked with the tracking result of each tracked face. For each detected face, if it does not correspond to any tracked face, then it is considered to be a new incoming face and its tracking model will be trained accordingly. Because faces sometimes may occlude with each other, it is necessary to design an occlusion detection and recovery method. If one tracked face is confirmed to be occluded by some other tracked face, a suitable occlusion recovery process should be performed so that this occluded face can resume its tracking whenever it separates with the occluding face again.

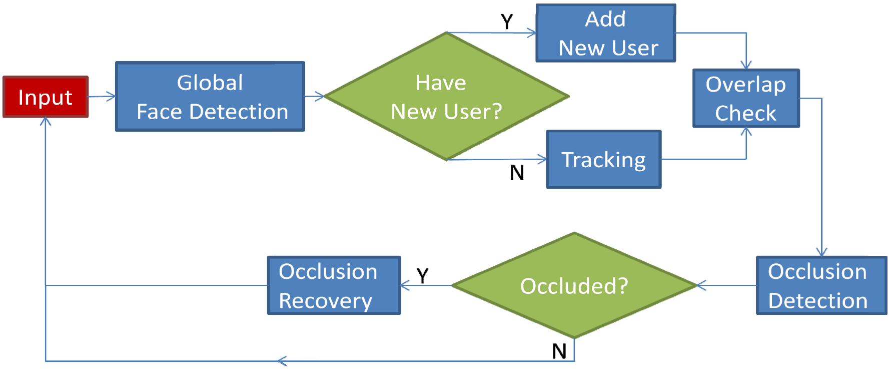


Fig. 2 Flow chart of multiple-face tracking system

3.4 Face Occlusion Detection

The purpose of face occlusion detection is to examine whether any two tracked faces overlap with each other or not. If two faces indeed overlap with each other, then it needs to further decide which face is the occluding one and which is the occluded one. With this decision, when the two overlapped faces separates with each other again in near subsequent frames, the occluded face is able to not be mistakenly recognized as a new face, and instead the system can resume its tracking with the preserved tracking model constructed just before it is occluded. The detailed occlusion detection process is described in the following:

1. When the overlapped portion of two tracked faces is between 5% and 30% of their average size, the two corresponding tracking face models will stop updating.

2. When the overlapped portion of two tracked faces is larger than 30% of their average size, it indicates a significant occlusion has occurred, and the face with the lower/higher tracking score is respectively regarded as the occluded/ occluding face.

After confirming the occurrence of significant face occlusion, the tracking model of the occluding face will resume to update, and the tracking model of the occluded face will be preserved without updating.

3.5 Recovery of Occluded Face Tracking

When a significant occlusion has occurred, in each subsequent frame we will decide whether there are enough moving skin-color pixels in any neighborhood candidate region of the occluded face. In order to detect moving skin-color pixels, temporal frame difference (TFD) and skin color detection (SCD) are used. TFD is to find the pixels having significant color change between either two consecutive frames or two N-frame apart frames. If one pixel has a significant color change, its TFD value is set to 1, otherwise it is set to 0. Although TFD can effectively detect the pixels of the moving objects, it unfortunately also contains many ghost pixels. Ghost pixels are the background pixels but occluded by a moving object in the previous frame. SCD is to detect the skin color from the image. Since both TFD and SCD have their own respective drawbacks, they are further intersected to produce a moving skin-color image and only the skin-color-like moving pixels can be reserved. If there are enough moving skin-color pixels, the preserved tracking model of the occluded face OA will be used to determine the reappearance of the occluded face. If a face OA has reappeared, then this face will be resumed its tracking from its reappeared location. The occlusion recovery is designed as follows:

1. Around the occluding face, a set of neighboring candidate regions (NCR) is defined and each region has the same size as the OA before it is occluded. Since it is extremely abnormal for an occluded face to reappear in front of the occluding face, hence, NCR can only be defined at the left side, right side and upper side of the occluding face as shown in Fig. 3, in which the green frame, the blue frame and the white frames denote respectively the occluding face, the occluded face, and several NCRs.
2. Both TFD and SCD are operated on the current image, and the two results are further intersected to generate a moving skin-color image.
3. Filter out the NCRs which have no enough moving skin-color pixels.
4. Compare with the preserved tracking model of OA, select the NCR having the highest tracking score from the remaining NCRs.
5. If the tracking score of the selected NCR is high enough, set the NCR as the reappeared location and resume tracking of OA.

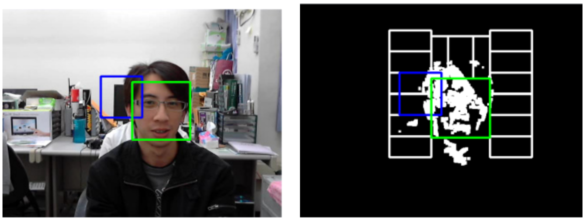

(a) Original image (b) Candidate region

Fig. 3 Illustration of candidate region

4 Experiment Results

For experiment, a self-designed database is constructed, and it contains (1) a single-person face database, and (2) a multiple-person face database. The single-person face database is taken by 7 persons with no occlusion by other objects and contains two types of angle constraints. In the first type, each person can rotate his/her face between -90° and 90°, and it contains 23 videos with 3577 frames in total. In the second type, each person can rotate his/her face without any angle constraint, and it has 22 videos with in total 3716 frames. In the multiple-person face tracking database, there are four persons involved and it contains 12 videos taken also under different places with in total 3716 frames. Among the 12 videos, there are 33 sequences of front-face occlusions and 30 sequences of side-face occlusions. Each occlusion contains a sequence of images formed by three sections, namely, approaching occlusion, under occlusion and gradual separation. Fig. 4 shows a few examples of both the single-face tracking and multiple-face tracking databases.

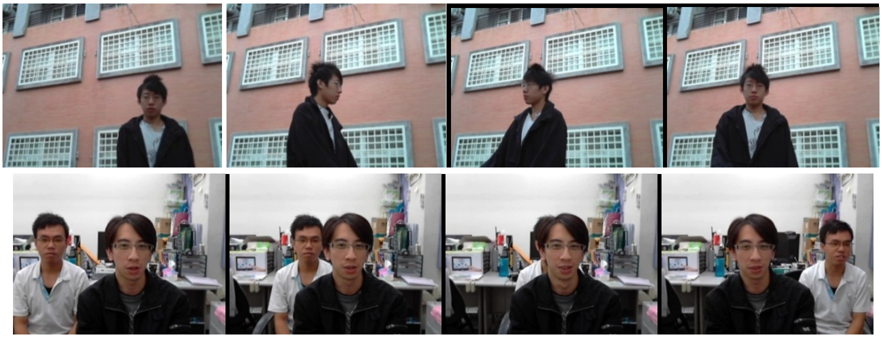


Fig. 4 Examples of face tracking database

Two kinds of experiments were performed, the first one is to track a single person without occlusion, and the second one is to track two persons at the same time and some of them are occluded with each other. Each face in one frame will have two corresponding rectangles (A and B). Rectangle A is shown in red and is set manually to serve as the groundtruth position of one face, and rectangle B is shown in green and is obtained by a tracking program automatically. The overlap rate (OR) of A and B is then computed as $\mathrm{OR}(\mathrm{A},\mathrm{B}) = \frac{|\mathrm{A} \cap \mathrm{B}|}{0.5 \times |\mathrm{A} \cup \mathrm{B}|}$. Only when the overlap rate is larger than 0.7, the tracking of this face in one frame is regarded as correct.

For performance comparison, besides the proposed method, both Camshift and Kalman Filter methods are also performed on the same test data. Table 1 shows the correct rates of the three methods for the single face tracking, and it clearly shows the proposed method achieves the best performance. Since the used Kalman Filter comes from OpenCV and has no stop condition, it performs poorly and will not be used in the following experiments.

Table 1 Single face tracking

data method	Constraint single face tracking	Unconstraint single face tracking
Camshift	87.92%	82.62%
Kalman Filter	68.44%	62%
Our method	94.13%	87.92%

The second experiment is to analyze the performance of multiple-face tracking. It consists of two kinds of occlusions, front-face occlusion and side-face occlusion. Front-face/side-face occlusion denotes a tracked face is in a front/side orientation before entering occlusion. Table 2 and Table 3 display respectively the recovery performance of front-face occlusion and side-face occlusion and they show (1) the proposed method has the ability to resume 70%~80% tracking even the tracked faces are once occluded by other objects and reappear after, and (2) the proposed method perform better than the Camshift method in dealing with the occlusion problem.

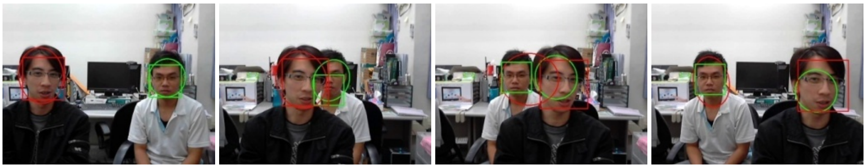
Table 2 Experiment results of front-face occlusion recovery

data method	Occlusion number	recovery number
Camshift	33	18
Our method	33	26

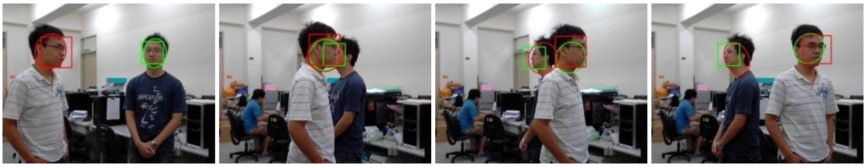
Table 3 Experiment results of side-face occlusion recovery

data method	Occlusion number	recovery number
Camshift	63	33
Our method	63	46

A few tracked results of multiple-face tracking are displayed in Fig. 5. The red/green color denote respectively the two tracked faces which are located at the left/right side in the frame before occlusion. The rectangle and the oval denote the tracked results obtained respectively by the proposed method and the Camshift algorithm. Fig. 5(a) and Fig. 5(b) are the occlusion recovery results when the occluded person is respectively in a front face and a side-face condition. It can be seen that the proposed method not only can resume tracking after occlusion but also can keep the same color on the same person. However, CamShift is not able to track the same object after it separates with the occluding object.



(a) Tracking recovery of front-face occlusion



(b) Tracking recovery of side-face occlusion

Fig. 5 Tracking recovery of face occlusion

5 Conclusion

This paper has associated face tracking process with face detection, and proposed a multiple-face tracking method. The proposed method can improve the commonly-encountered drifting error in the traditional tracking methods. In the single face tracking, the face detection result is used to compensate the tracking result, therefore, the accuracy of tracked face can be effectively improved. Also, in the multiple-face tracking, through the proposed occlusion recovery mechnism, pretty good recovery result can be obtained even when a tracked face once is occluded by other faces.

Acknowledgments This work is supported by the Ministry of Science and Technology with grant no. MOST 103-2221-E-216-009 and MOST 103-2622-E-216-003-CC3.

References

1. Elgammal, A., Harwood, D., Davis, L.: Non-parametric model for background subtraction. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 751–767. Springer, Heidelberg (2000)
2. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence **17**, 185–203 (1981)
3. Wu, M., Peng, X.: Spatio-temporal context for codebook-based dynamic background subtraction. Int. J. Electron. Commun. (2009)
4. Zhu, S., Ma, K.K.: A new diamond search algorithm for fast block-matching motion estimation. IEEE Trans. Image Processing **9**, 287–290 (2000)
5. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach toward Feature Space Analysis. IEEE Trans. Pattern Analysis and Machine Intelligence **24**(5), 603–619 (2002)
6. Nummiaro, K., Koller-Meier, E., Gool, L.V.: An Adaptive Color-Based Particle Filter. Image and Vision Computing **21**, 99–110 (2003)
7. Breitenstein, M., Reichlin, F., Leibe, B., Koller-Meier, E., Gool, L.V.: Robust tracking-by-detection using a detector confidence particle filter. In: ICCV (2009)
8. Babenko, B., Yang, M.-H., Belongie, S.: Visual tracking with online multiple instance learning. In: CVPR (2009)
9. Zeisl, B., Leistner, C., Saari, A., Bischof, H.: Online semi-supervised multiple-instance boosting. IEEE Conference on Computer Vision and Pattern Recognition (2010)

Regions Labeling in Outdoor Scene Images

Kyawt Kyawt Htay and Nyein Aye

Abstract Outdoor scene analysis is a complex problem for both image processing and pattern recognition domains. This paper proposes an approach for labeling regions in outdoor scene images. The basic idea of this approach is to label local image regions into semantic objects such as tree, sky and road etc. There are four phases in the approach: segmentation, feature extraction, region labeling and merging. Firstly, modified Marker-Controlled Watershed (MCWS) algorithm proposes for segmented regions generation. And then, color feature extracted from segmented regions are given as input to 3-layer Artificial Neural Network (ANN) classifier for labeling. Finally, region merging is performed if the two regions are adjacent with the same color values. The proposed method is test on our real scene image dataset which are collected from our university campus.

Keywords Region labeling · Segmentation · ANN classifier · Region-merging

1 Introduction

Scene understanding is a fundamental problem in robotics and perception environment. Most robotic scene understanding work has focused on outdoor scenes captured with laser scanners, where technologies have matured and enabled high-impact applications such as mapping and autonomous driving [1]. Autonomous robots capture visual information of scenes and train a classifier, predict the taversability of real scene images with the trained classifier. The promising classification results provide the basis for obstacles avoiding and path planning. Gi-Yeul Sung et al. [2] presented a terrain classification method based upon color and texture features,

---

K.K. Htay(✉)
University of Computer Studies, Mandalay, Myanmar
e-mail: kyawtkyawthtay2014@gmail.com

N. Aye
Head of the Department of Hardware, University of Computer Studies,
Mandalay, Myanmar
e-mail: nyeinaye@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_26

they extracted texture features with discrete wavelet transform coefficients mean and energy values. Furthermore, they adopted spatial coordinates where a terrain is located in the image as additional features. Color histograms are a popular method of representing the image [3]. Moreover, image segmentation is one of the most complex stages in image analysis. It becomes essential for subsequent image description and recognition tasks. The problem consists of partitioning an image into its constituent regions, objects or labels [4]. The level of division depends on the specific problem being solved. This partition is accomplished in such a way that the pixels belonging to homogeneous regions, regarding to one or more features (i.e., brightness, texture or color), share the same label, and regions of pixels with significantly different features have different labels. There must be considered four objectives for developing an efficient generalized segmentation algorithm [5]: continuous closed contours, non-over segmentation, independence of threshold setting and short computation time. In particular, the over-segmentation error, which occurs when a single semantic object is divided into several regions, is a tendency of some segmentation methods like watershed [6]. Therefore, some subsequent region merging process is needed to improve the segmentation results.

2 Related Works

Many scene classification systems needs to be sure which objects are in the image before classifying the scene. The common method of doing this is to segment the objects in the image and then identify the low-level features and semantic concepts. There are two methods to segment images, one is block-based and one is region-based [7] [8]. The block-based method simply segments the image into several rectangles. The region-based method also segments the image, but the objects are more meaningful to human eye. The objective of a scene understanding system consists of recognizing and localizing the significant imaged objects in the scene and identifying the relevant object relationships. Consequently, a system must perform segmentation, region characterization and labeling processes. Gao-Hua, Iaeng and Zhao Chun-xia, Zhang Hao-feng [9], proposed a scene classification method. First, texture features are extracted in gray channel. Then, color moments are computed in all three color channels of RGB color space. The parameters of GMM are estimated by training these labeled samples. Finally, new scene images are classified by the trained GMM model. Vogel and Schiele [10] modeled the semantic content of an image and used this model to classify local image regions into semantic concept classes. Model-based systems rely upon the configuration of the scene components. Luo and Stephen [11] proposed a model-based approach to detecting sky. This approach consists of color classification, region extraction and physics motivated sky signature validation.

In this paper, our contribution is modified Marker-Controlled Watershed (MCWS) algorithm is proposed and region merging is performed on classifier output image. Therefore, our approach will reduce computational complexity and time than others pixel-based and clustering-based segmentation approaches. Our work will focus on the problem of image regions labeling to classify every region of a given image into one of several pre-defined classes.

3 Watershed Transform

The advantages of watershed transformation are that it is simple, instinctive knowledge, and can be parallelized [12]. The main drawback of this method is the over-segmentation due to the presence of many local minima. To decrease the effect of severe over-segmentation, marker-controlled watershed transformation is used. Separating touching objects in an image is one of the more difficult image processing operations. The watershed transform is often applied to this problem. The watershed transform finds "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low.

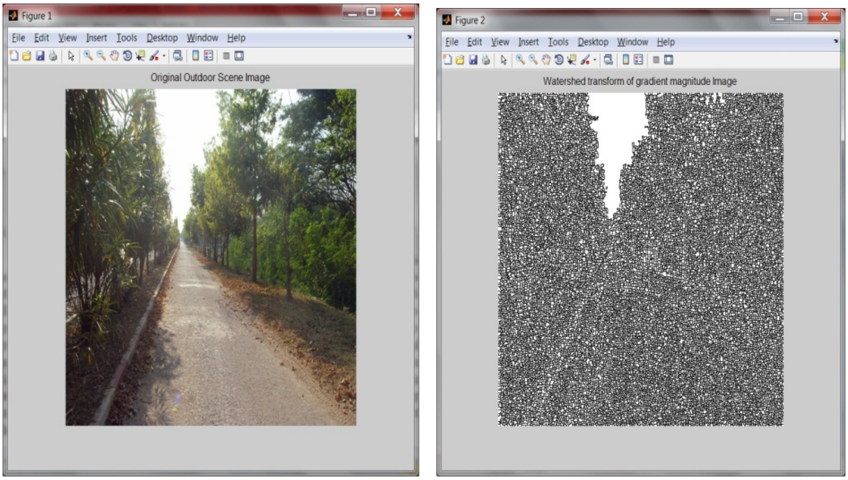


Figure 1
File Edit View Insert Tools Desktop Window Help
Original Outdoor Scene Image
Figure 2
File Edit View Insert Tools Desktop Window Help
Watershed transform of gradient magnitude Image

(a) (b)

Fig. 1 (a) Original Image (b) Result of Standard Watershed Algorithm

The above described figure 1. (a) and (b) show original outdoor scene image and segmented result of standard watershed algorithm. The standard watershed algorithm produces 23833 segmented regions. Therefore, we need to reduce the number of segmented regions by using marker controlled approach.

3.1 Modified Marker-Controlled Watershed (MCWS) Algorithm

This paper proposes modified Marker-Controlled Watershed (MCWS) algorithm for segmented regions generation. There are three steps in the propose algorithm: pre-processing, watershed transform and post-processing. In the proposed modified MCWS algorithm, a variety of procedures could be applied to find the foreground markers which must be connected blobs of pixels inside each of the foreground objects. The method that is used to find foreground markers is morphological techniques called "opening-by-reconstruction" and "closing-by-reconstruction" to "clean" up image. These operations will create flat maxima inside each object that can be located using imregionalmax. Background markers are computed in the cleaned-up image by using a thresholding operation. And then compute watershed transform of the segmentation function.

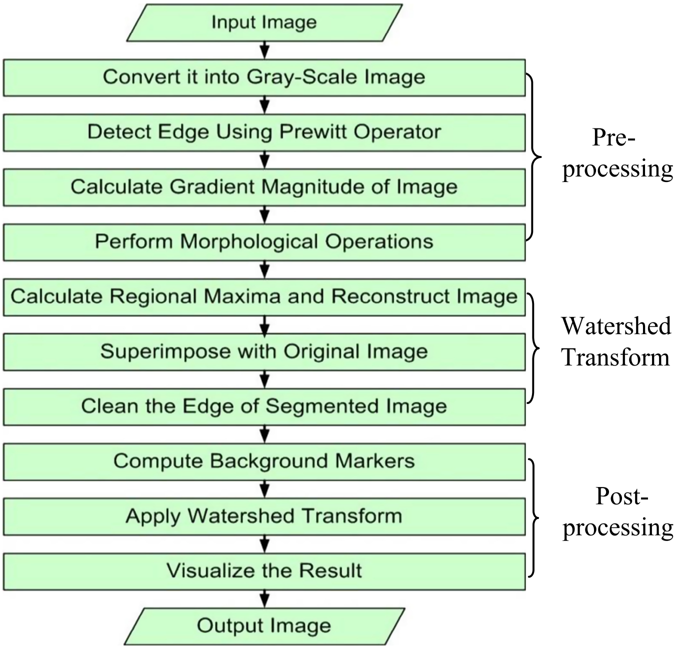

Input Image
Convert it into Gray-Scale Image
Detect Edge Using Prewitt Operator
Calculate Gradient Magnitude of Image
Perform Morphological Operations
Pre-processing
Calculate Regional Maxima and Reconstruct Image
Superimpose with Original Image
Clean the Edge of Segmented Image
Watershed Transform
Compute Background Markers
Apply Watershed Transform
Visualize the Result
Post-processing
Output Image

Fig. 2 Flow chart of proposed modified MCWS algorithm

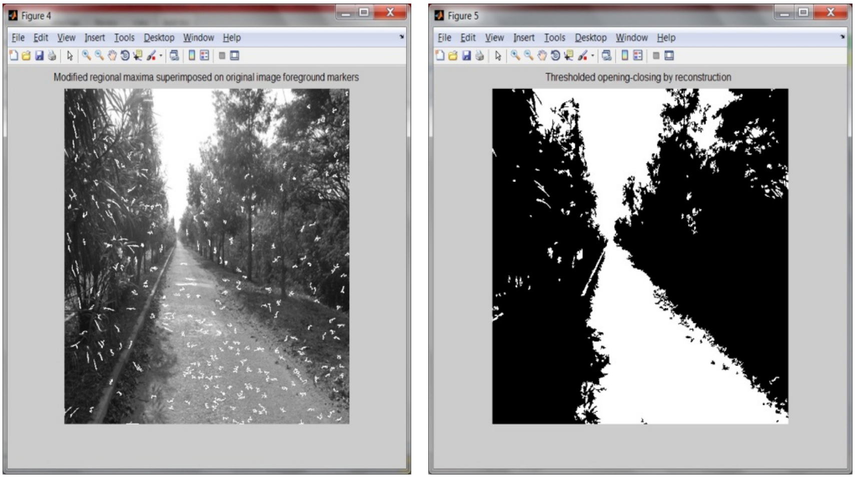


Figure 4
File Edit View Insert Tools Desktop Window Help
Modified regional maxima superimposed on original image foreground markers
Figure 5
File Edit View Insert Tools Desktop Window Help
Thresholded opening-closing by reconstruction

(a) (b)

Fig. 3 (a) Foreground Marker by using Morphological Operations (b) Background Marker by using Thresholding Method

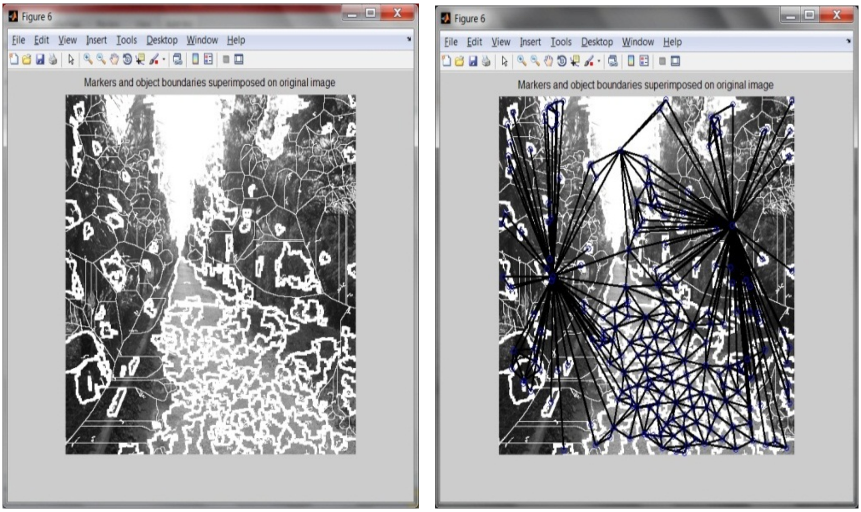


Figure 6
File Edit View Insert Tools Desktop Window Help
Markers and object boundaries superimposed on original image
Figure 6
File Edit View Insert Tools Desktop Window Help
Markers and object boundaries superimposed on original image

(a) (b)

Fig. 4 (a) Result of Modified MCWS Algorithm, (b) Region Adjacency Graph

The figure 4 (a) shows the result of our proposed modified MCWS algorithm. Our proposed modified MCWS algorithm can reduce over segmentation problem of standard watershed algorithm by using marker controlled. The number of

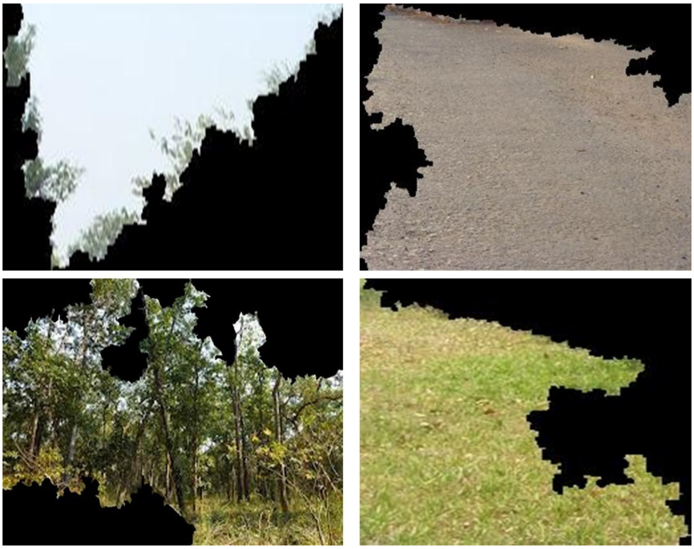


Fig. 5 Regions Dataset

segmented regions produce is 215 regions. Therefore, the proposed modified MCWS algorithm can reduce 23618 segmented regions than standard watershed algorithm. The segmented regions can give useful information for later steps such as feature extraction and region-labeling. The segmented regions are collected as a region dataset is shown in figure 5. The first row represent for sky and road. The second row represent for tree and grass etc.

4 Feature Extraction

An important step in classification is that of feature extraction. Here, 768 color features are extracted from each region for classification purposes. Color is one of the most effective, simplest and widely used low level visual features. Color has been thought of as an important property of objects. It can be used for segmentation and classification.

4.1 Color Histogram

In the analysis of color images, the description of image regions has mainly been performed using color histograms. The color histogram is obtained by counting the number of times each color occurs in the image array. Color histogram H for a given image is defined as a vector

$$\mathbf{H} = \{H[0], H[1], \ldots, H[n]\} \tag{1}$$

Where i represent the color in color histogram and H[i] represent the number of pixels of color i in the image, and N is the number of bins used in color histogram. To extract color features, we need to represent the color in a suitable space. This system uses RGB color space to extract color features. The number of bins used in color histogram is 256+256+256=768 bins for RGB channels. Therefore, the 768 feature sets are given as input to 3-layer ANN classifier for regions labeling.

5 Regions Labeling

5.1 Classifier

In this system, 3-layer Artificial Neural Network (ANN) classifier is used to classify local image regions. Using the feature set described above, a neural network with 768 inputs and 5 label outputs has been trained and optimized. Conjugate gradient descent optimization was used and the optimal network architecture was found to have 24 nodes in the hidden layer.

Training. In the training step, weighting parameters are computed through off-line training by supervised learning using error back-propagation. Once an input pattern is applied to the network, the weights are adjusted iteratively by the back-propagation method, which minimizes the mean square error between the actual and desired output. We set the learning rate to 0.3 and, the training process is iterated until the mean square error is smaller than 0.1. In order to train the network, 768 color features are used in RGB color spaces.

Running. In the running step, the 3-layer ANN classifier is performed by loading the network parameters computed from the training results. The classified result can then be computed using feature vectors extracted from an input image. Neural network inputs are made up of weighting parameters and feature vectors. The output of the network is the classified region type to which the input region belongs.

5.2 Region Adjacency Graph (RAG)

Region adjacency graph (RAG) which is a usual data structure used to represent region neighborhood relations in a segmented image [13]. RAG is a weighted undirected graph G={V,E,W}, where each node represents a region of the over segmented image and each edge is a symmetric dissimilarity function between adjacent regions.

Region Merging. Region merging is a post-processing technique that merges adjacent regions with the same color values. In the proposed region merging approach, there are two predicates.

- The first predicate is minimum pair for each region.
- The second one is color homogeneity distribution.

From RAG, minimum edge pairs for each region are discovered. Color homogeneity distribution can be obtained from 3-layer ANN classifier output. Only if the region pair can satisfy these two predicates, the region pair is actually merged. If not, this region pair is ignored.

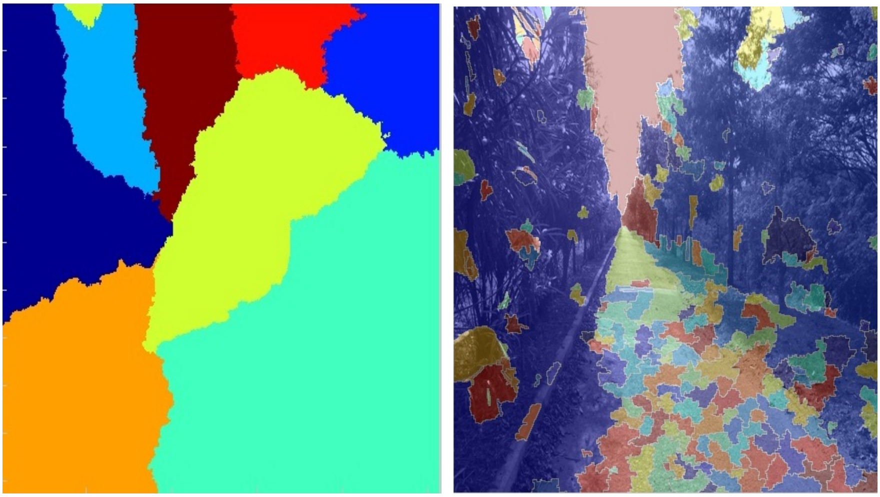
6 Results and Discussion

The proposed method is applied on our real scene images dataset was collected from our university campus. All of the images were taken during daytime to give a realistic view of the real environment. Environmental conditions during capture were dry and good atmospheric visibility. All of the images collected are stored in JPEG (.jpg) format with 24-bit color depth and the size of image is resize into 512x512 pixels. The 768 color features extracted for every region are presented to 3-layer ANN classifier and produces five output labels, which include sky, road, building and tree etc. And then, to reduce the number of labeled regions, region merging is performed if the two regions are adjacent with the same color values.

As a limitation, our system cannot accept under segmented images. However, the system can accept over segmented image. So the segmented image must not be under segmented image for later feature extraction and region-labeling steps.

Some segmentation methods like N-cut at which is a well-known segmentation method but like all methods it only works on specific datasets. The problem of the N-cut algorithm is to determine the best N value which is the number of segments in an image. Determining a fixed N-value is not very reliable so N-cut gives satisfactory result only for specific images with specific N-values.

When our proposed modified MCWS is compared with N-cut segmentation, the N-cut segmentation gives under segmented image. The proposed method can gives satisfy result than N-cut and traditional watershed algorithm.



(a) (b)

Fig. 6 (a) Result of N-cut Segmentation and (b) Result of our proposed modified MCWS

The above figure 6. (a) Shows the result of N-cut segmentation on the input scene image with N-value is 8. The result image faces the problem of under segmentation for the input N-value. N-cut segmentation is not used in the system because it produces under segmented image. In (b) shows the result using our proposed modified MCWS algorithm. The proposed modified MCWS algorithm gives suitable over segmented image for later feature extraction and region-labeling steps. Therefore our system uses the proposed modified MCWS algorithm for segmentation.

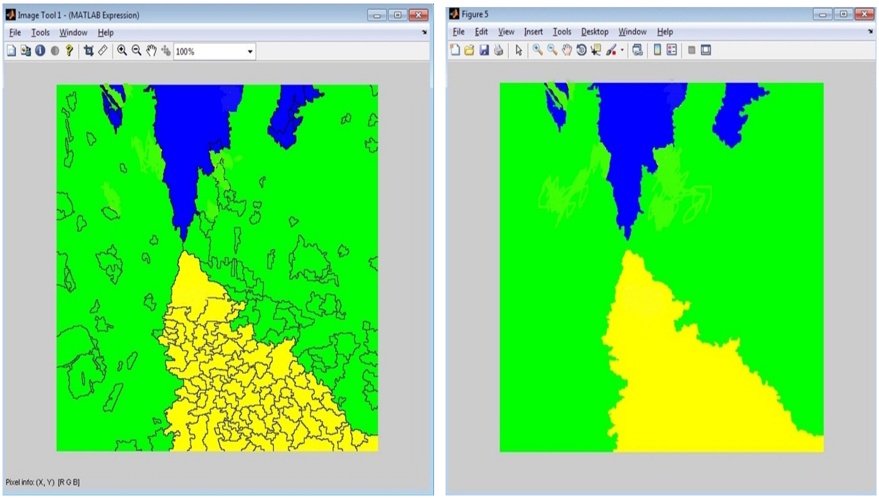


Image Tool 1 - (MATLAB Expression)
File Tools Window Help
100%
Pixel info: (X, Y) [R G B]
Figure 5
File Edit View Insert Tools Desktop Window Help

(a) (b)

Fig. 7 (a) 3-layer ANN Classifier Output Image (b) Result of Region Merging Image

🟦 Sky 🟨 Road 🟩 Tree

The above described image (a) and (b) show the result of 3-layer ANN classifier output and the result of after region-merging. The blue color represents for sky, yellow color represent for road and green color represent for tree etc. After region merging, we can reduce the number of labeled regions 215 into 7 regions. Therefore, our proposed regions labeling method can produce meaningful regions efficiently.

7 Conclusion

In this paper, modified Marker-Controlled Watershed (MCWS) algorithm is proposed for segmented regions generation. The proposed modified MCWS algorithm can produce suitable over segmented image than N-cut and standard watershed algorithm. The segmented image is used for later feature extraction and region-labeling steps. Color histogram is used in RGB space to extract color features. The extracted 768 color features are given as input to 3-layer ANN classifier to label segmented regions. In the classifier output image, many labeled regions are adjacent with the same color values. Therefore, region merging is performed to produce meaningful labeled regions. For future work, the proposed method can be test with various segmentation algorithms and low-level features that are used for region labeling.

References

1. Xiong, X., Munoz, D., Bagnell, J., Hebert, M.: 3-D scene analysis via sequenced predictions over points and regions. In: Proc. of ICRA (2011)
2. Sung, G.-Y., Kwak, D.-M., Lyou, J.: Neural Network Based Terrain Classification Using Wavelet Features. Intelligent Robot System **59**, 269–281 (2010)
3. Chapelle, O., Haffner, P., Vapnik, V.N.: Support vector machines for histogram-based image classification. IEEE Transactions on Neural Networks **10**(5), 1055–1064 (1999)
4. Gonzalez, R.C., Woods, R.: Digital Image Processing, 2nd edn. Prentice Hall, Englewood Cliffs (2002)
5. Ho, S.Y., Lee, K.Z.: Design and analysis of an efficient evolutionary image segmentation algorithm. J. VLSI Signal Process **35**, 29–42 (2003)
6. Haris, K., et al.: Hybrid Image Segmentation Using Watersheds and Fast Region Merging. IEEE Trans. Image Process **7**(12), 1684–1699 (1998)
7. Ko, C., Lee, H.S., Byun, H.: Image retrieval using flexible image sub-blocks. In: Proceedings of the 2000 ACM Symposium on Applied Computing 2000, pp. 574–578, March 2000
8. Szummer, M., Picard, R.W.: Indoor-outdoor image classification. In: Proceedings of IEEE International Workshop on Content-Based Access of Image and Video Databases, Bombay, India, pp. 42–51 (1998)
9. Gao, H., IAENG, Zhao C.-X., Zhang H.-F.: Visual features fusion for scene images classification. In: Proceedings of the International Multi-Conference of Engineers and Computer Scientists, IMECS 2012, Hong Kong, vol. II, March 14–16, 2012
10. Vogel, J., Schiele, B.: Semantic modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision (2004)
11. Luo, J., Etz, S.P.: A physical model-based approach to detecting sky in photographic images. IEEE Transactions of Image Processing **11**(3), 201–212 (2002)
12. Beucher, S., Meyer, F.: The morphological approach to segmentation: the watershed transform. In: Dougherty, E.R. (ed.) Mathematical Morphology in Image Processing, vol. 12, pp. 433–481. Marcel Dekker, New York (1993)
13. Sonka, M., et al.: Image processing, analysis and machine vision, 2nd edn. PWS (1999)

Segmentation of Video Image Using Changing Detection and Block Based Approach

Mie Mie Tin and Mie Mie Khin

Abstract Segmentation of object from video image sequence is very important in many aspects of multimedia application. We present an approach to discover and segment foreground object(s) in video. Our approach is to identify 8x8 blocks in any frame and compute a series of binary partitions among those blocks and compare each frame of blocks finding change of all frames in images. Our approach is automatically segmentation of object on the persistent foreground regions of interest. We apply our method to challenge standard videos, and show competitive or better results than the state-of-the-art.

Keywords Video image processing · Image segmentation · Changing detection · Block processing

1 Introduction

Image processing has a wide range of practical applications, such as car assembly, robot vision, airport security, object recognition, criminal investigation, and medical image analysis for robotic doctor. Image segmentation is an image processing technique to separate mutually exclusive homogeneous regions of interest of an image. Segmenting objects in an image play a fundamental role in the field of image processing and image analysis. Segmenting objects in an image is a very difficult and challenging task due to huge number of objects and the variations among the objects in terms of color, intensity and locations. [12]

M.M. Tin
University of Computer Studies, Mandalay, Myanmar
e-mail: Miemietin1983@gmail.com

M.M. Khin(✉)
Myanmar Institute of Information Technology, Mandalay, Myanmar
e-mail: Miemie.khin9@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_27

Image segmentation subdivides the image into its constituent objects. It has many approaches. Pixel based segmentation, which uses only pixel appearance to assign a label to a pixel [10, 11]. Region based segmentation, the algorithm detects for valid segments at each scale further segments extracted at various scale are integrated to get the final result [13]. In graph-based labeling methods [3, 14], a global energy functions defined depending on both appearance and image gradients. Mean shift based segmentation is been widely-used for the segmentation which includes two steps, filtering the original image data in feature space and clustering of the filtered data points. [5]

Finding moving objects in image sequences is one of the most important tasks in computer vision. Segmentation of moving object in image sequence plays an important role in image sequence processing and analysis. Our paper is to describe the boundaries of all moving and static objects occurring in video. In general, objects are spatially unified and characterized by locally smooth motion trajectories. Therefore, they occupy regions within each video frame. Image segmentation base algorithms are many processes, such as watershed transform and color segmentation on each frame to separate a frame into many homogeneous regions. By combining motion information derived with motion estimation, optical flow, or frame difference, regions with motion vectors different from the global motion are merged as foreground regions.

Our approach identify block base algorithm for compare frame of video images sequence. The objectives of this paper are to get and identify video image frames and moving objects from the standard video images. Our approach will be find reasonable image frames and improve analysis quality for video object segmentation.

2 Related Work

A. Criminisi group presented GeoS, a new algorithm for the efficient segmentation of n-dimensional image and video data. They proposed new method to get efficient geodesic distance computation was used to propose a set of spatially smooth, contrast-sensitive segmentation hypotheses. Their Advantages include: i) computational efficiency with high segmentation accuracy; ii) the ability to estimate an approximation to the posterior over segmentations; iii) the ability to handle generally complex energy models. Comparison with max-flow indicates up to 60 time greater computational efficiency as well as greater memory efficiency. GeoS is validated quantitatively and qualitatively by thorough comparative experiments on existing and novel ground-truth data. Numerous results on interactive and automatic segmentation of photographs, video and volumetric medical image data were presented.[4]

B. Qi et. al purposed a hierarchical differential GME oriented to object segmentation. A combination of 3-step search and motion vector prediction was proposed for initial estimate. Two robust estimators were also proposed: to estimate global motion in the first frame and to reject outliers using object information. Subjective and objective results show that the proposed method was more robust and faster than the reference methods. [17]

D. Zhang et.al. Purposed to achieve high compression performance, the objects segmented may not be semantically meaningful to human observers. They also describe many segmentation techniques have been proposed in the literature, fully automatic segmentation tools for general applications are currently not achievable. Their paper provides a review of the important and challenging area of segmentation of moving objects. They described common approaches including temporal segmentation, spatial segmentation and the combination of temporal-spatial segmentation. [18]

M. Grundmann et. al presented an efficient and scalable technique for spatiotemporal segmentation of long video sequences using a hierarchical graph-based algorithm. Their hierarchical approach was generated high quality segmentations, which was temporally coherent with stable region boundaries, and allows subsequent applications to choose from varying levels of granularity. They proposed two novel approaches to improve the scalability of their technique: (a) a parallel out-of-core algorithm that can process volumes much larger than an in-core algorithm, and (b) a clip-based processing algorithm that divides the video into overlapping clips in time, and segments them successively while enforcing consistency. They were demonstrated hierarchical segmentations on video shots as long as 40 seconds, and even support a streaming mode for arbitrarily long videos, albeit without the ability to process them hierarchically. [8]

N. Friedman described that background subtraction the current image from a time- averaged background image would leave only non- stationary objects. The basic idea of his paper was that they could classify each pixel using a model of how that pixel looks when it was part of different classes. Their approach was automatically updates the mixture component for each class according to likelihood of membership; hence slow-moving objects are handled perfectly. They identified and eliminated shadows much more effectively than other techniques such as threshold. They also expected to result in significant improvements in vehicle identification and tracking. [7]

Thi Thi Zin et. al proposed a new framework for detecting suspicious people and suspicious objects by using an integration of probability based background modeling, high level motion features, embedded Markov chain and the first passage time probabilities. In their integrated framework, they were first established a probability based multiple backgrounds modeling technique to detect moving object. Then the high level motion features of objects of interest such as velocity of end distance between the semantic (objects) in the scene were extracted. [16]

W. Brendel presented an approach to unsupervised segmentation of moving and static objects occurring in a video. They used a low-level segmentation to extract regions in all frames and also find similar regions was transitively matched and clustered across the video. They formulate a new circular dynamic-time warping (CDTW) algorithm that generalizes DTW to closed contours, without compromising the optimality and low complexity of DTW. [1]

X. Bai et. al described an interactive framework for soft segmentation and matting of natural images and videos . Their proposed technique was based on the optimal, linear time, computation of weighted geodesic distances to the user-provided scribbles, from which the whole data is automatically segmented. The weights were based on spatial and/or temporal gradients, without explicit optical flow or any advanced and often computationally expensive feature detectors. [2]

3 Approach

Combine change detection approach was using Thoma and Bierling to carry out object segmentation. M Hotter et. al were used change detection based segmentation algorithm. This research is used to find segmentation objects based their approach.[9,15] This approach has three step. Firstly the threshold operation is performed on measurement window using mean absolute frame difference and second part is two dimensional median filter for smooth the boundaries of changing region. Finally, is to eliminate small isolated regions .These three step using until the new threshold until system is stable. The segmentation of moving object is achieve by separating the covered and uncover background from the change detection mask. Blocks subtraction approach can eliminate these uncover background from each video frame and compare each block approach can reduce memory usage of calculation time.

The recognition of moving objects is important in many tasks. Other researcher approaches to this problem can be broadly divided into two classes: 2D algorithms which apply when the scene can be approximated by a flat surface and/or when the camera is only undergoing rotations and zooms, and 3D algorithms which work well only when significant depth variations are present in the scene and the camera is translating. We describe a unified approach to handling moving object detection in video sequence. Our approach is based on a stratification of the moving object detection problem into scenarios which gradually increase in their complexity. We present a set of techniques that match the above stratification. We show that techniques using examples from standard image sequences. In our approach using motion estimation process for video frame. Figure 1 shows that the process of propose object segmentation for video image.

The general representation of transformation is

$$[x \quad y \quad 1] = [u \quad v \quad 1]\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix}$$

If $w = a_{11}, a_{12}, a_{21}, a_{22}, a_{31}, a_{32},$ is the vector of the motion parameters, the following system of equations must be solved for each object in the scene $Aw = b;$

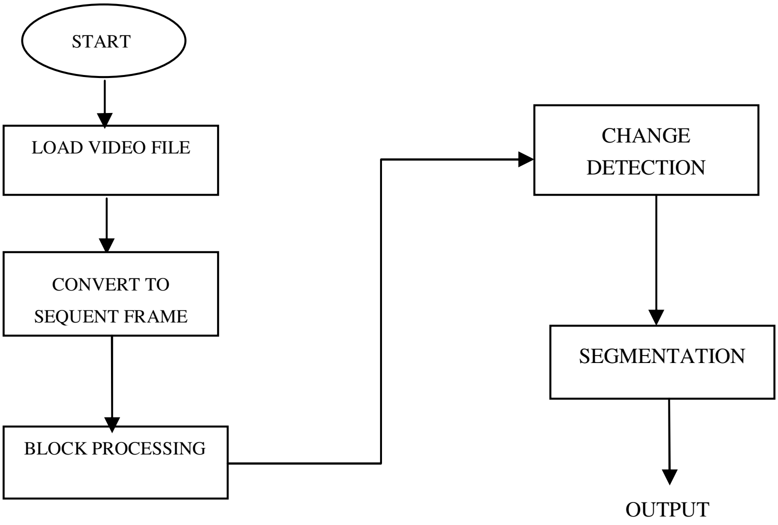


Fig. 1 Block diagram purpose image segmentation process

The solution to the above over determined set of equations may be obtained by use of a least square method, or alternately by the robust least median of squares technique.

In each video frame, we consider the block consideration method will be use for each image. In our block consideration method consists of two parts, namely flag encoding and block data encoding. Flag encoding records the location and color information of the block targets, and block data encoding records get block data and their boundary information. Video file from frame, first of all, we divide the 8x8 images frame from into four pieces of 2x2 square blocks.

Block Data subtraction, we do two steps. First step is getting the block data if first frame and second frame. Second step is calculating the block subtraction in each block by block. In get block data, we get the same color block size from calculate block size. Block data will get xNode + row, yNode + clm and same color block size.

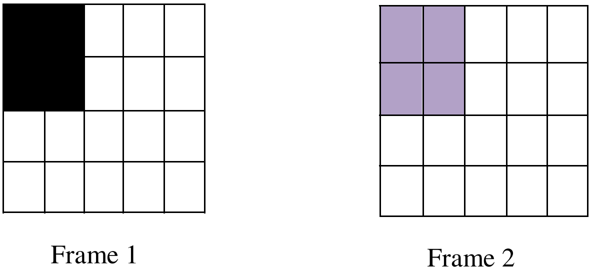


Fig. 2 Block Data for two video Images Frame

In calculate each block, we calculating the pixel value with pixel by pixel based row order.

In this calculation, the main intended area is first row and first column of block and last row and last column of image. Because we need to know the beginning of block that remark as a start region of block, and we need to know the end of the image that to stop the indexing.

If first frame blocks and second frame block subtraction is near zero these two blocks are same and otherwise are difference. This difference blocks have different motion of objects. We consider that difference block and extract these new objects using 2D Motion Estimation process. Our approach considers that camera is still, and there is no light changing and no shadows. It is based on change detection mask here is not only generated from the frame difference of current frame and previous frame but also from frame difference between current frame and background frame. Figure 3 show that sample comparison idea between video frames.

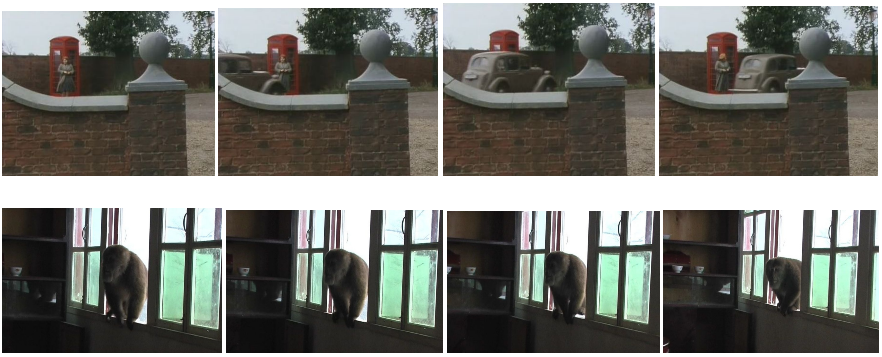


Fig. 3 Test Images form Marple17 and Monkey 5 Movie Databases

4 Conclusion

This paper purposes object segmentation based on block each frame of video image. This algorithm can generate segmentation result with low computation complexities and memory usage. This purpose algorithm has limitations, that is light changing and shadows in each frame. Furthermore, this purposed algorithm is not automatic for all processes. This algorithm is only designed for moving objects segmentation and the foreground object cannot still be long time. Furthermore we will compare other moving objects segmentation methods using our analysis results.

References

1. Brendel, W., Todorovic, S.: Video Object Segmentation by Tracking Regions. IEEE int'l Conf., Computer Vision (ICCV), Kyoto, Japan (2009)

2. Bai, X., Sapiro, G.: A Geodesic Framework for Fast Interactive Image and Video Segmentation and Matting. IEEE 11th International Conference on Computer Vision, ICCV 2007. IEEE (2007)
3. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects. In: Proceeding of international Conference on Computer Vision (ICCV-01), vol. 2, pp. 105–112 (2001)
4. Criminisi, A., Sharp, T., Blake, A.: GeoS: geodesic image segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 99–112. Springer, Heidelberg (2008)
5. Comanicu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. In: IEEE Trans. Pattern Anal. Machine Intell, vol. 24, pp. 603–619 (2002)
6. Dong, Y., Maqueak, D.S.: Unsupervised Segmentation of Color-Texture Regions in Images and Video (2001)
7. Friedman, N., Russell, S.: Image segmentation in video sequences: a probabilistic apprach. In: Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc. (1997)
8. Grundmann, M., Kwatra, V., Han, M., Essa, I.: Efficient Hierarchical Graph-Based Video Segmentation. Circuits, Systems and Signal Processing **20**(2), 143–183 (2001)
9. Hotter, M., Thoma, R.: Image Segmentation Based on Object Oriented Mapping Parameter Estimation. Signal Processing **15**, 315–334 (1988)
10. Hlanning, T., Pisinger, G.: A pixel-based segmentation algorithm of color images by n-level-fitting. In: The 5th IASTED International Conference on Computer Graphics and Imaging (2002)
11. Das, M., Manmatha, R., Riseman, E.M.: Indexing flower patent images using domain knowledge. IEEE Intelligent Systems **14**, 24–33 (1999)
12. Hasan, M.M., Sharmeen, S., Rahman, M., Ali, M.A., Kabir, M.H.: Block based image segmentation. In: Das, V.V., Stephen, J. (eds.) CNC 2012. LNICST, vol. 108, pp. 15–24. Springer, Heidelberg (2012)
13. Mukhopadhyay, S., Chanda, B.: Multi scale morphological segmentation of gray-scale image. IEEE transactions on Image Processing **12**, 533–549 (2003)
14. Nilsback, M.E, Zisserman, A.: Automated flower classification over a large number of classes. In: The proceeding of Sixth International Conference on Computer Vision, Graphics and Image Processing, pp. 722–729 (2008)
15. Thoma, R., Bierling, M.: Motion Compensating Interpolation Considerating covered and Uncovered Background. Signal Processing: Image Communication **1**, 191–212 (1989)
16. Zin, T.T., Tin, P., Hama, H., Toriu, T.: An Integrated Framework for Detecting Suspicious Behaviors in Video Surveillance. IIS 13-067, pp. 41–48
17. Qi, B., Amer, A.: Robust and fast global motion oriented to video object Segmentation. In: IEEE International Conference on Image Processing, ICIP 2005, vol. 1. IEEE (2005)
18. Zhang, D., Lu, G.: Segmentation of Moving Objects in Image Sequence: A Review. Circuits, Systems and Signal Processing **20**(2), 143–183 (2001)

Part IV
IPv6 Operational Technologies and Network Applications

A Study of the IPv6 Service Availability Measurement

Shian-Shyong Tseng, Ai-Chin Lu, Ching-Heng Ku, Sheng-Wei Kuo and Geng-Da Tsai

Abstract While confronting the global IPv4 address exhaustion, it is important and crucial for the entire Internet environment to smoothly upgrade to the next generation Internet Protocol, IPv6. Based on the statistics of the user availability of Google, the global usage of the IPv6 is still very low. As we knew, the service availability is one of the key factors affecting user availability, how to handle the IPv6 service availability measurement becomes a critical topic for the IPv6 upgrade.

In previous researches, two points of views for estimating the IPv6 service availability have been proposed. One is to calculate the ratio of all ccTLD IPv6-enabled services to all ccTLD services, and the other is to calculate the ratio of all IPv6-enabled services in the top k popular services to k. In the former method, the importance of each service is treated as equal and the related data need to be provided by its ccTLD. In the latter method, the top k popular services representing the all services are used to calculate the IPv6 service availability and the related data could be obtained by certain public web sites. In this study, we proposed a hybrid method to analyze the service availability using the global top one million popular web sites and top 500 ccTLD's popular web sites from Alexa.

In this study, four criteria have been defined and experiments have been made to verify our hypothesis.

Keywords IPv6 readiness · Service availability

S.-S. Tseng
Department of Applied Informatics and Multimedia, Asia University, Taichung, Taiwan
e-mail: sstseng@asia.edu.tw

A.-C. Lu · C.-H. Ku(✉) · S.-W. Kuo · G.-D. Tsai
Taiwan Network Information Center, Taipei, Taiwan
e-mail: {aclu,chku,cwkuo,dar}@twnic.net.tw

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_28

1 Introduction

Network measurement [1] is a measurement activity of network characteristics. It measures the network features, including network topology, network routing, network traffic, the network anomalies, network performance bottlenecks, etc. Generally, the collected measurement information can be used to improve the network operation and management.

In order to measure different network characteristics, the so-called Network Measurement Infrastructure was proposed [2]. It used multiple measuring points in different locations on the network to obtain the measurement results from each measurement point to infer the status of network environment.

Some distributed measurement architectures have been proposed [3], where the common features of these architectures rely on a few (e.g., dozens) of distributed measurement nodes, to measure large-scale network. Besides, the different type of the distributed measurement architecture [4] based on point to point type to construct a large-scale measurement system was also proposed. In the large-scale network measurement, data acquisition can be divided into direct access to the actual data and indirect estimation of the data in order to increase measurement accuracy.

Facing the depletion of IPv4 address, how to find a feasible approach which can smoothly and steadily upgrade from the IPv4 network to IPv6 network environment is a globally concerned issue.

Based on the statistics of the user availability of Google, the global usage of the IPv6 is still very low. As we knew, the service availability is one of the key factors affecting user availability, how to handle the IPv6 service availability measurement becomes a critical topic for the IPv6 upgrade.

In this study, four criteria, such as representativeness, popularity, accessibility, and coverage, have been defined and experiments have been made to verify our hypothesis. The hypothesis is that the ratio of the IPv6 web service availability for popular web services is at least higher than 5 times of the overall web services for any ccTLD in 2015. The gap will be slightly decreasing every year. The hypothesis can be used to predict the IPv6 service availability of a certain ccTLD in the next few years.

2 Related Works

The global organization, the Internet Corporation for Assigned Names and Numbers (IANA), allocated the last five IPv4 blocks to Regional Internet Registries (RIRs) on February 3, 2011[5]. The regional organization, Asia-Pacific Network Information Center (APNIC) managing the IP address in the Asia-Pacific region announced that the IPv4 address is exhausted and will stop the normal allocation of IPv4 addresses for Asia-Pacific countries on 15 April, 2011[6].

The Network Working Group of the Global Internet Engineering Task Force, IETF, pass through the RFC 5211 in July 2008 to announce an Internet Transition

Plan [7], which proposed the stage and schedule planning for the smooth transition of IPv4 networks to IPv6 network as a global reference basis for network deployment.

Until the end of April 2012, there are 184 countries/economies around the world have been allocated of the IPv6 address [8]. It means that many countries have paid attention to the development of IPv6 networks. Besides, many developed countries also give the priority to the IPv6 network upgrade in the government to promote the development of the IPv6 industry and the private enterprise.

There are some IPv6 development surveys in the world. The Internet Association Japan [9] and the European Commission [10] had started the IPv6 penetration survey to understand IPv6 development. In addition, the international IPv6 Status Survey [11] provides five measurement indicators. S. S. Tseng, et al. [12] proposed an IPv6 readiness model with seven criteria as the promotion indicators of IPv6 readiness. The IPv6 Taiwan Directory [13] shows the list of IPv6-enabled web sites. A service-based IPv6 readiness architecture, including five measurement criteria from the perspectives of service, software, and core network, is shown in Fig. 1.

Services	• **Ratio of public network Services on the IPv6 Web, E-mail, DNS, FTP**
Software	• **Ratio of the supporting IPv6 in the Operation System** • **Ratio of the supporting IPv6 in Service Software**
Core Network	• **Ratio of the supporting IPv6 in the network equipment for the access** • **Ratio of the supporting IPv6 in the network equipment for the management**

Fig. 1 Measurement Model of IPv6 Readiness with five measurement criteria

We assume that the popular services can represent the coverage of the service usage in the user availability. In this study, a hypothesis for IPv6 service availability is proposed. The hypothesis is that the ratio of the IPv6 web service availability for popular web services is at least higher than 5 times of the overall web services for any ccTLD in 2015. The global top one million web sites and ccTLD's top 500 popular web sites from Alexa are used to estimate the service availability of ccTLD. The statistical results successfully show our hypothesis for ccTLD's IPv6 service availability is true.

3 Methods for Service Availability

In this study, two points of views for the IPv6 service availability are proposed. One is to calculate the ratio of all ccTLD IPv6-enabled services to all ccTLD services, and the other is to calculate the ratio of all IPv6-enabled services in top k popular services to k, where top k web services are ranked by Alexa (www.alexa.com).

In the former method, the importance of all services is treated as equal and the data need be provided by every ccTLD. This method focuses on the domestic services and does not consider the access frequency of services.

In the latter method, top k popular services representing the most important services and used to calculate the IPv6 service availability by considering the access frequency of services, where the data could be obtained from Alexa[14].

Alexa publishes the popular websites by collecting and ranking over 1,000GB web sites' usage information. It is currently owned by the largest number of URL, ranking the most detailed information publishing site. Alexa website ranking mainly in two ways: ranking and classification rankings. Ranking also called absolute ranking, ie a specific site in which there website ranking. Alexa published every three months the new site ranking. This ranking is based on the number of links user (Users Reach) and the number of page views (Page Views geometric mean) three-month cumulative.[14]

Alexa ranks sites based primarily on tracking a sample set of internet traffic [14] — users of its toolbar for the Internet Explorer, Firefox and Google Chrome web browsers.[15][16] It also allows the user to rate the site and view links to external, relevant sites. This caused some controversy over how representative Alexa's user base was of typical Internet behavior, [17] especially for less-visited sites.[16]

In this study, we proposed four criteria, such as representativeness, popularity, accessibility, and coverage, for the evaluation of the measurement method for the IPv6 service availability. The degree of these criteria can be shown in the Table 1.

Table 1 The degrees of four criteria for the evaluation of the measurement method of the IPv6 service availability

Criteria	Definition	Degree
(A) Representativeness	Measured services are sufficient to represent the overall number of services.	1. Low 2. Moderate 3. High
(B) Popularity	Measured services are considered from the different importance by the access frequency.	1. Non 2. World wide 3. Domestic 4. Both
(C) Accessibility	Data can be automatically acquired.	1. Low 2. Moderate 3. High
(D) Coverage	The ratio of the number of measured services to overall number of services.	1. Low 2. Moderate 3. High

The higher representativeness the measured services are, the more suitable the overall number of services represent. The popularity considers the access frequency of measured services that are considered from the different importance measured in the domestic or that in the world wide. The accessibility can be determined whether the measured data can be automatically acquired or not. In other words, the data is more difficult to be acquired, so the accessibility is low. The coverage means the ratio of the number of measured services to overall number of services. The larger the ratio means that the higher the coverage. Based on the four criteria, the comparison of two points of views for the IPv6 service availability is shown in Table 2.

Table 2 Comparison of two points of views for the IPv6 service availability

Methods	Definition	Criteria				Comparison
		A	B	C	D	
(I) Equal Weighted method in ccTLD	Calculate the ratio of all ccTLD IPv6-enabled services to all ccTLD services.	3	1	1	3	1. Data shall be provided by ccTLD. 2. The importance of each service is treated as equal.
(II) Top k method in Alexa	Calculate the ratio of all IPv6-enabled services in top k popular services to k.	2	2	3	2	1. The high access frequency of services are considered. 2. Consider the world wide access frequency.

4 Statistical Analysis for IPv6 Service Availability Measurement

In this study, we use the above criteria to analyze the measurement methods of the IPv6 service availability and proposed the hypothesis for the statistics of the IPv6 service availability. Based upon the collected service data from Taiwan, New Zealand, Singapore, and Vietnam, we want to verify the proposed hypothesis. Furthermore, the trend of the IPv6 service availability can also be estimated in the future.

In this study, we proposed a hypothesis that the ratio of the IPv6 web service availability for popular web services is at least higher than 5 times of the overall web services for any ccTLD.

The results of four measurement methods for Taiwan are shown in Table 3. In the first method, we calculate the ratio of all .tw IPv6-enabled services to all .tw services. In the second method, we calculate the ratio of the number of IPv6-enabled .tw web services to the number of .tw web services in top 1 million of Alexa. In the third method, we calculate the ratio of the number of IPv6-enabled web services to 500, top 500 web services of Taiwan ranked by Alexa.

Table 3 Result of four measurement methods of the IPv6 service availability for Taiwan

Method for measured data	# of measured Services (A)	# of IPv6-enabled services (B)	Ratio of IPv6 web service availability (B)/(A)
(I).tw web services	396,486	7,074	1.78%
(II-1) .tw web service in top 1 million of Alexa	4,694	364	7.75%
(II-2) Taiwan web service in top 500 of Taiwan ranked by Alexa	500	64	12.8%

As mentioned above, we knew that the accessibility for the data acquisition of the method (I) is low, because not all ccTLD registries are willing to provide users' data. On the opposite, the top 1 million service data in Alexa has cover at least 200 ccTLD services. Therefore, we want to estimate the IPv6 service availability based on the public data of Alexa web site. According to our previous knowledge and experience on the IPv6 readiness measurement, we make the following hypothesis.

Hypothesis: The ratio of the IPv6 web service availability for popular web services is at least higher than X times of the overall web services for any ccTLD.

$$\mathrm{X} = 5 * \left\{1 - \frac{1}{10} * (n - 2015)\right\},\ \ 2015 \le n < 2020, n\ is\ the\ year$$

The popular web services in the above hypothesis are acquired from top 500 of a country ranked by Alexa. We collect the ccTLD service data of May 2015 from Taiwan, New Zealand, Singapore, and Vietnam to observe the proposed hypothesis. The definition of three measurement methods is shown in Table 4.

Table 4 Definition of four measurement methods of the IPv6 service availability

Methods	Definition
(I)	% of IPv6-enabled ccTLD web services
(II)	% of IPv6-enabled ccTLD web services of the ccTLD services in Top 1 million of Alexa
(III)	% of IPv6-enabled web services in top 500 of a country ranked by Alexa

The Comparison of four ccTLD in different measurement methods of the IPv6 service availability is shown in Table 5.

Table 5 Comparison of four ccTLD in different measurement methods of the IPv6 service availability

Methods	Taiwan (.tw)	New Zealand(.nz)	Singapore (.sg)	Vietnam (.vn)
(I)	7,074/396,486 =1.78%	8,202/640,788 =1.28%	3,147/172,899 =1.82%	3,072/278,364 =1.10%
(II)	364/4,694=7.75%	55/1,315=4.18%	204/1,323=15.42%	73/3,554=2.05%
(III)	64/500=12.80%	40/500=8.00%	64/500=12.80%	43/500=8.60%
(III)/(I)	7.2	6.3	7.0	7.8

Based on the four ccTLD data in the above, we observe that the ratio of the IPv6 web service availability for popular web services is at least 6 times of the overall web services. This statistical result shows our hypothesis for ccTLD's IPv6 service availability is true.

Besides, we hope to know the relation of the IPv6 service availability and the IPv6 user availability in a country. We use APNIC IPv6 user availability to compare the IPv6 service availability at measurement method (II) in top 1 million of Alexa, shown in Table 6.

Table 6 Comparison of APNIC IPv6 user availability and proposed IPv6 service availability at measurement method (II)

Country (Country code)	IPv6 User Availability from APNIC	(B/A) IPv6 Service Availability of ccTLD in top 1 million of Alexa	(A) # of ccTLD in top 1 million of Alexa	(B) # of IPv6-enabled services in (A)
Belgium (.BE)	38.66%	4.56%	2563	117
Germany (.de)	19.53%	12.56%	28193	3540
United States of America (.us)	17.90%	6.18%	2607	161
Norway (.no)	11.01%	6.71%	2431	163
Switzerland (.ch)	10.14%	5.08%	2856	145
Czech Republic (.cz)	9.21%	16.17%	4719	763
Japan (.jp)	9.03%	1.37%	21796	298
Greece (.gr)	8.99%	13.05%	5550	724
Romania (.ro)	8.54%	4.09%	3518	144
France (.fr)	5.57%	6.50%	11948	777
Austria (.at)	4.27%	5.62%	2649	149
Taiwan (.tw)	0.71%	7.75%	4694	364
China (.cn)	0.54%	0.53%	11261	60

Among the top 1 million web services of Alexa, there are around 0.43 million web services belonging to 246 ccTLDs. Because the number of web services collected by Alexa is too small for certain ccTLD, the ccTLD that has more than 2,000 ccTLD web services in Alexa can be chosen for the comparison in this study.

We observe that there is no positive correlation between the IPv6 user availability measured by APNIC and the IPv6 service availability based on Alexa top 1 million web services for a country.

5 Conclusion

Facing the depletion of IPv4 address, how to find a feasible approach which can smoothly and steadily upgrade from the IPv4 network to IPv6 network environment is a globally concerned issue. Based on the statistics of the user availability of Google, the global usage of the IPv6 is still very low. As we knew, the service availability is one of the key factors affecting user availability, how to handle the IPv6 service availability measurement becomes a critical topic for the IPv6 upgrade.

We assume that the popular services can represent the coverage of the service usage in the user availability. The global top one million and top 500 popular web sites from Alexa are used to estimate the service availability of ccTLD.

In this study, we proposed a hypothesis and four criteria, such as representativeness, popularity, accessibility, and coverage, for the evaluation of the measurement method for the IPv6 service availability. The hypothesis is that the ratio of the IPv6 web service availability for popular web services is at least higher than X times of the overall web services for any ccTLD. The X is 5 in Year 2015 and will be decreased 10% every year up to 2020. The popular web services in the above hypothesis are acquired from top 500 of a country ranked by Alexa. Based upon the collected service data from Taiwan, New Zealand, Singapore, and Vietnam, we verified the proposed hypothesis.

Based on the four ccTLD data in the above, we observe that the ratio of the IPv6 web service availability for popular web services is at least 6 times of the overall web services. The statistical results successfully show that our hypothesis for ccTLD's IPv6 service availability is true. The trend of the IPv6 service availability will be estimated based on our methods in the future.

Acknowledgement This paper is partially sponsored by Ministry of Transportation and Communications, Republic of China, under Grant 1040212coco, and Taiwan Network Information Center (TWNIC).

References

1. Caceres, R., Duffield, N., Feldmann, A., Friedmann, J.D., Greenberg, A., Greer, R., Johnson, T., Kalmanek, C.R., Krishnamurthy, B., Lavelle, D., Mishra, P.P., Rexford, J., Ramakrishnan, K.K., True, F.D., vander Memle, J.E.: Measurement and Analysis of IP Network Usage and Behavior. IEEE Communications Magazine **38**(5), 144–151 (2000)
2. Wijata, Y.I., Niehaus, D., Frost, V.S.: A scalable agent-based network measurement infrastructure. IEEE Communications Magazine **38**(9), 174–183 (2000)
3. Internet Measurement Infrastructure. http://www.caida.org/analysis/performance/measinfra
4. Li, C.-F.: P2P-Based Programmable Network Measurement Infrastructure. Thesis, Nationa Tsing Hua University (2004)
5. IANA: Global Policy for the Allocation of the Remaining IPv4 Address Space, February 3, 2011. http://www.icann.org/en/news/in-focus/global-addressing/remaining-ipv4
6. APNIC: Key Turning Point in Asia Pacific IPv4 Exhaustion. http://www.apnic.net/__data/assets/pdf_file/0018/33246/Key-Turning-Point-in-Asia-Pacific-IPv4-Exhaustion_English.pdf
7. IETF RFC 5211: An Internet Transition Plan. http://64.170.98.42/html/rfc5211
8. Global IPv6 address statistics. http://trace.twnic.net.tw/ipstats/statsipv6.php
9. Measurement of IPv6 readiness, Internet Association Japan. http://v6metric.jp/en/index.html
10. European Commission: IPv6 Deployment Monitoring, December 14, 2010. http://ipv6-ghent.fi-week.eu/files/2010/12/1335-Rob-Smets-v2.pdf
11. IPv6 status survey. http://www.mrp.net/IPv6_Survey.html
12. Tseng, S.S., Lu, A.C., Ku, C.H., Tsai, G.D.: A Study on Measurement Model of IPv6 Readiness. In: Proceedings of 2010 Cross-Strait Conference on Information Science and Technology, CSCIST 2010, pp. 647–651. Yanshan University, Qinhuangdao, July 9–11, 2010
13. IPv6 Taiwan Directory. http://v6directory.twnic.net.tw
14. Alexa Internet, Wikipedia. http://en.wikipedia.org/wiki/Alexa_Internet
15. Technology: How and Why We Crawl the Web. Alexa. Archived from the original on, April 2, 2014
16. Davis, H.: Google Advertising Tools: Cashing in with AdSense, Adwords, and the Google APIs, p. 12. O'Reilly Media (2006). ISBN 978-0-596-10108-4
17. Croll, A., Power, S.: Complete Web Monitoring: Watching Your Visitors, Performance, Communities, and Competitors, p. 38. O'Reilly Media (2009). ISBN 978-0-596-15513-1

Automated Home Agent Configuration for Mobile IPv6

Wuttipon Noopetch and Sinchai Kamolphiwong

Abstract Mobility support in IPv6 is an extension IPv6 protocol. It provides for mobile IPv6 node which relocates to another network location. Mobility support in IPv6 can keep connection between mobile node and the global node on Internet after mobile node changes a network location. However when mobile node moves and changes a network location, it cannot send data to another node for keep connection. Because mobility support in IPv6 protocol needs to send packet via home network of mobile node, resulting in a triangle route between two nodes that build a cost of mobility support in IPv6 protocol. So we proposed a solution for solving the cost of the mobility support in IPv6 and get the benefit for keeping connection between two nodes.

Keywords Mobile IPv6 · Multiple home agents

1 Introduction

Mobility support in IPv6 (MIPv6) [1] solves a movement of mobile device problem. When a mobile device uses a MIPv6, it can move to another network and keep connection. This is a strong point of a MIPv6 for continuously working of a mobile device. However the most connections of mobile devices are often very short and move much less frequently than the average connection period. So this is a cause of MIPv6 problem. Because communications are routed via old network, it increases packet delays and the possibility of loss. Moreover a problem with fragmentation packet for encapsulate tunnel with first network. Unfortunately this may lead user to disable MIPv6 for avoiding that cost. However if the mobile device is at home, there is no cost. The proposed solution makes every possible locations can be the home. This paper concentrates on the first step that how to

---

W. Noopetch(✉) · S. Kamolphiwong
Department of Computer Engineering, Faculty of Engineering,
Centre of Network Research, Prince of Songkla University, Hat Yai 90112, Thailand
e-mail: wuttipon@icloud.com, ksinchai@coe.psu.ac.th

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_29

automate the operation of MIPv6 (base on IPv6). The remainder for using and evaluating the above operation for MIPv6 is the further work.

2 Background Information

2.1 Mobility Support in IPv6

IP addresses are dual purpose. One use is to identify end points involved in communications. They are also used to provide the location of an end point for packet delivery. Because of the second use, when a node moves its IP address must change. Because of the first use, that change causes existing communications to fail. Mobile IP was created to overcome this problem, it allows communications to continue after a node has moved, and so obtained a new address. Mobile IP (MIP) is defined in RFC 3220 [2].

A Mobile Node (MN) has as its Home Address (HoA) the same value as its Care-of Address (CoA) when it is on its home link. Packets from Correspondent Nodes (CNs) are sent to the MN using its HoA as the destination address. The reach of the home network is using the conventional Internet routing mechanism, where they are delivered to the MN as to any other node. The MN replies directly to the CN using its HoA as the source address. When the MN moves from the home link to a foreign link, it will obtain a CoA using normal IPv6 mechanisms, such as stateless or statefull auto-configuration. The MN will then send a Binding Update (BU) message to its Home Agent (HA) so the HA can update its binding cache and remember the association between the MN’s HoA and its current CoA. The HA will reply to this binding registration with a Binding Acknowledgement (BA) message to inform the MN of the result of its update. Once this has completed the MN can resume communications with the CN. In Mobile IPv6 it has two choices for how this is accomplished.

The first method is bidirectional tunneling shown in figure 1, a packet from the CN will be sent to the home network first. Then HA will intercept the packet and will tunnel it to the MN. The HA uses proxy Neighbor Discovery [3] to intercept the IPv6 packet, that arrives for the MN on the home link. To reply to the CN, packets are tunneled between MN and HA (reverse tunnel) so they can be routed to the CN from the home link. The advantage of bidirectional tunneling method is that the CN does not need to participate at all, the MN’s CoA is not known to it.

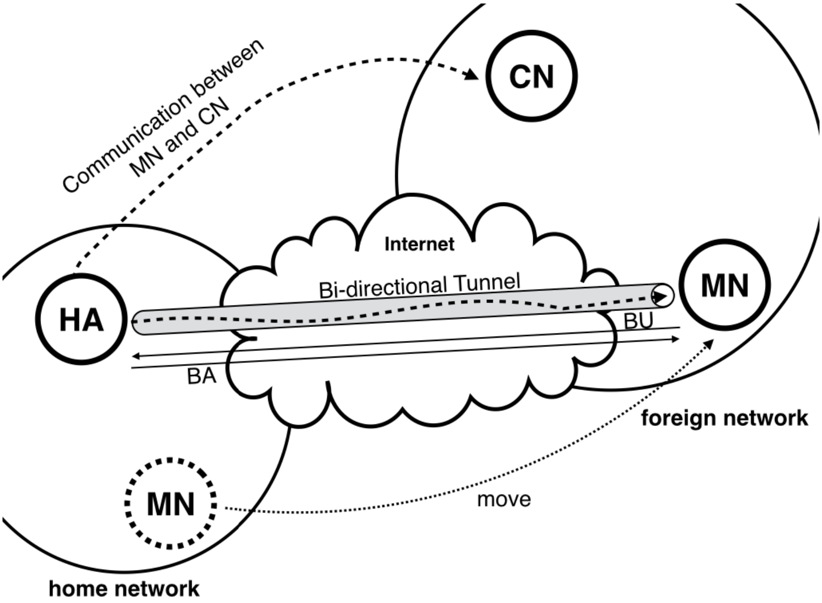


Fig. 1 MIPv6 Operation.

2.2 *Route Optimization*

When using MIPv6, and with support of the CN, Route Optimization (RO) can be used. With this, the MN sends a BU to the CN. So RO can solve triangle route problem from standard MIIPv6. However before MN and CN establishing bi-direction tunnels to communicate each other, the CN needs a method to verify HoA and CoA that MN is an owner. So the process of MIPv6 for verification is called Return Routability (RR) [4]. When the CN needs to send a packet from the transport layer to any IPv6 destination, it first checks the binding cache for the destination address. If it is found, the CN will use an IPv6 routing header [5] to directing the packet to the HoA via the CoA. The MN sends a HoA information to CN by using a new option of MIPv6 is a Destination Options Header as the Home Address option (HAO). So the CN can communicate directly to the MN by using a Routing Header Type 2 (RTHDR2). This results in the packet travel directly to the MN at its CoA, where it then “routes” the packet to itself at its alternate address. Advantages to using RO include a shorter transmission path between MN and CN, and less congestion at the HA.

Unfortunately the RO can solve triangle route problem from standard MIPv6, but it causes a cost of communicate between MN and CN. Because all packets sent from CN need to check binding table stacks before sending to MN, this is the cost of the packet overhead and cache lookup of CN.

2.3 Multiple Home Agents

There is some support for multiple HAs in MIPv6, intended to support reconfiguration of the home network. Such a reconfiguration could lead to a MN that is away from home being unable to locate its HA, as it does not know the new HA IP address. The MN can use Dynamic Home Agent Address Discovery (DHAAD) to discover the IP addresses of HAs. The MN sends a Home Agent Address Discovery message to one of the HA's using a well known anycast address [6]. The anycast address is constructed from the prefix of the network on which the HA is expected, which is the same as the prefix of the HoA of the MN, and a defined constant. The DHAAD procedure causes the HA that receives the Discovery message to reply to the MN informing it of the addresses of HAs on the selected prefix. This allows a MN to regain contact with its HA. This mechanism does not provide any configuration, the MN must already know its HoA, and thus the prefix of its HoA, the MN must also already have a relationship with the HA in order to successfully use the HA address information discovered.

3 Problem Statement

The MN stays at home network and connects to the CN node. MIPv6 can keep connection between MN and CN when MN moves to the foreign network. That is a benefit of standard MIPv6. MN needs to send all data (and BU) via HA. This is a cost of Standard MIPv6. When MN has a new connection with any IPv6 node, the MN cannot connect to IPv6 node.

The RO creates connection direct to CN when MN leaves at home network, but some popular server maybe not support this function. It causes a large BU table for maintain all of MN. So this is a cost of the popular server that it supports mobility function.

3.1 Triangle Route on Mobile IP

- Packet delay added by triangle routing causes problem with real time application for example Video Conference, Multicast Video.
- The performance of Transmission Control Protocol (TCP) depends upon the Round Trip Time (RTT) between the client and server. A longer RTT means slower maximum transmit rate. Further more hops means a bigger chance of packet loss, and retransmission.
- Packet loss because MNs send packet to CN via the HA. It uses more bandwidth and cause traffic congestion at the HA.
- Tunnel between HA and MN reduces MTU of that link, which may reduce path MTU between CN and MN (usually will). Which means more packets request for connections that use long packet (TCP) or more fragmentation for large UDP (eg. DNS with DNSSEC). Both load to increased congestions and packet loses.

3.2 A Large Binding Cache on CN

When an MN is away from home it uses route optimization to contact a CN. The CN must remember the CoA of all the MNs in contact with it. That could make the binding cache of the CN large, especially if the CN is a popular server show in figure 2. So this is speculation as correctly MIPv6 is so little used and supported by few CNs that these is no current experiment .And might cause the CN to refusing to use route optimization, in which case an MN has only triangle routing and increasing delays, as a solution.

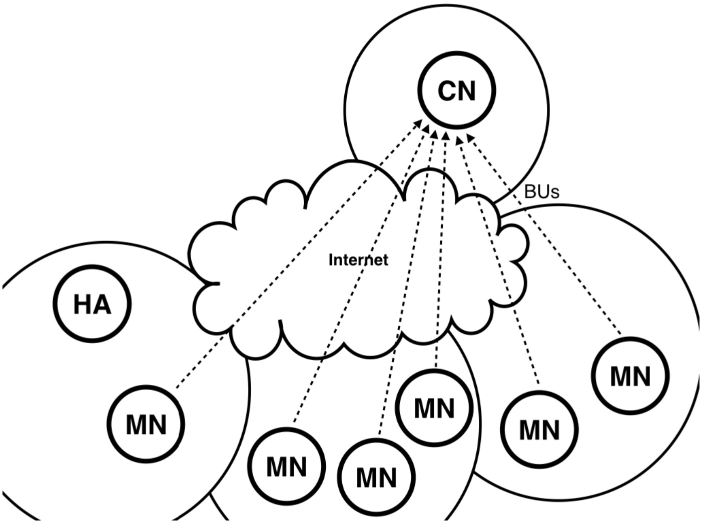


Fig. 2 Binding Cache of CN get Large.

4 Proposed Solution

We observe that there is no significant cost to using Mobile IP when the MN is at home, only when it is away from home do the triangle route, or route optimization costs appear. So a solution to the problem would be for the MN to always be at home. That is, whenever an MN moves its new location becomes a new home if possible. This requires the MN to associate with an HA in its new network. The MN will then have multiple HAs.

To do this the MN must first locate an HA on its new network, and then configure its association with the MN. Current implementations involve people in this process, for our scheme to work everything must be automated. This includes finding an HA, obtaining its agreement to serve as HA for this MN, and configuring the required security association between the HA and MN.

Once the MN is able to add new HAs as required it might eventually obtain many HAs. Eventually some of those will be serving no useful purpose and should be removed. We need to find a way to detect HAs that are no longer of any use so they can be deleted.

Not all networks will have an HA available for the MN to use. In that case Mobile IP will work as it has before, including the extra costs. But now the MN has a new problem, when it initiates communications it needs to use a home address, but it now has several available - which is the best for it to choose? We need to seek an answer to this to provide guidance to MN implement while any choice would be functional, an optimal choice would be better.

4.1 Automatic Home Agent Configuration

We proposed multiple HAs method for reduce cost of MIPv6. The MN can find new HA immediately when it moves to a new network. If the user needs to configure MIPv6 every time it moves. Then the user maybe revokes the connection about MIPv6. So we need a new protocol for this method automated. Figure 3 show a new protocol diagram we proposed.

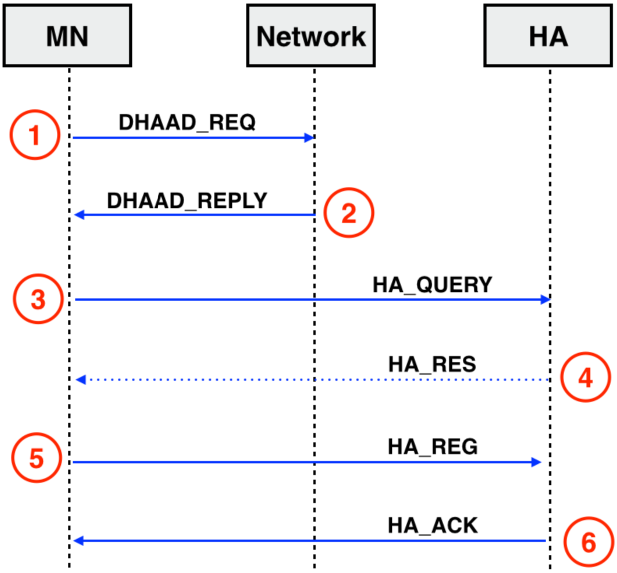


Fig. 3 Protocol Diagram.

Packet 1-2.

When MN moves to new network, it needs to find a new HA for a communication by use MIPv6 service. If this network using DHCPv6 then MN can get address and HA addresses from same time.

The packet DHAAD_REQ MN send Dynamic Home Agent Address Discovery (DHAAD) [1] request to a network. And then MN gets HAs list from DHAAD reply (DHAAD_REPLY packet) on a network. The standard of DHAAD protocol use for recover HA address when MN has been away from home and HA's has a new address. But our protocol design uses DHAAD to get a HAs address list for automatic configuration of new HA when it moves to new site.

Packet 3.

The purpose of HA_QUERY packet is to query a HA for MIPv6 service. Actually MN need to identify with HA. But the first time MN can not trust that a real HA. So MN send only public information to a HA. This information not secret and any body can read it. So the HA_QUERY packet contain with

- Local address (CoA): this address can send BU to old HAs and identification on new network.
- Primary HA: If the addresses are CGAs. It can verify MN.
- Nonce: is a 64 random bits for security issue. Because when HA reply back this packet. MN can check it and protection from bogus reply.
- Addition: If the MN has some kind of identification from some network. It should be included. Because it likely a certificate and some network allow MN to use MIPv6 service. But if MN has HA's public key from

But if HA has been authenticated as a router and uses Secure Neighbor Discovery (SEND) [7]. Then we have a certificate, a public key. So we can authenticate HA before packet 3.

Packet 4.

The answer of request from HA has tree kind (no/yes/maybe).

4a: If HA replies "no" to MN. The reply packet contains HA authorization but it is not essential. Because MN can ignore and choose other HA, a nonce needs to be included in a packet MN to check that packet form HA.

If HA replies another answer (yes/maybe). MN needs to proof an HA authorization in this packet. The method for proofing that likely verifies a router on Secure Neighbor Discovery (SEND) [7]. So if MN doesn't trust a chain of certificate. It can use a mechanism from SEND router verification for proof that.

4b: So if HA replies "yes" answer that information in HA_QUERY packet sufficiently for HA decided.

4c: HA replies "maybe" answer that it need more information from MN (inside HA_QUERY packet). And then MN can send a data need to be secret encrypt. Because MN has a public key from the certificate of HA (inside HA_RES packet) for encrypted data, it likely a replay HA_QUERY packet and additional data (eg. Paypal account, credit card number) with secret encrypt.

A HA not keep a state when reply HA_RES packet. If HA answer reply "no", MN ignore that HA and try new one. If answer reply "maybe", this a state likely a reply "no". Because MN can determine try different HA or send more information. Final answer (reply "yes") MN can get a several from HA's same time. So MN can pick a only one HA.

This packet contain with:

- Answer reply: (yes/no/maybe)
- Certificate: inside this has a public key from HA. The public key use for verify the signature of HA.
- Nonce:The MN know this packet replay from HA_QUERY packet.
- List of support shared-key security algorithms: The MN can select then (eg. AES, 3DES, Blowfish) and use it to encrypt secret data next packet.

Packet 5.

This packet is like a HA_QUERY packet, but it has an encrypted data block after MN getting a certificate from HA. And. this packet MN send to only one HA's. So this packet contain with

- Local address (CoA) : this address can send BU to old HAs and identification on new network.
- Primary HA: If the addresses are CGAs. It can verify MN.
- Nonce: is a 64 random bits for security issue. Because when HA reply back this packet. MN can check it and protection from bogus reply.
- Addition: If the MN has some kind of identification from some network. It should be included. Because it likely a certificate and some network allow MN to use MIPv6 service.
- Share secret key: When MN choose security algorithm from HA_RES packet, it uses shared secret key for encrypted packet between MN and HA.

Packet 6.

This packet is likely a HA_RES packet. If HA answers reply "no" (reject), HA needs to give a reason for MN.

5 Conclusion and Evaluation

When a mobile node is away from home, we have two costs from MIPv6. The solution to solving that problem is to make sure that the MN is always at home. And we make a multiple home agents for solving that problem. We need an automatic HAs configuration protocol for look through and establish new HA.

An implementation of an automatic HA's configuration protocol is working in progress. It will be optimized nearly for the rest of the project. Under a condition that HA will exist of each network and it will say "yes" at packet 4. A future work of a method to automatic HAs configuration is demonstrated to exist, functional and security. We can go on to testing operation of MIPv6 using multiple HA's of different home sites and then investigate the remaining issues.

Acknowledgment I would like to thank Mr. Robert Elz. He has devoted a lot of time to me and is always very patient with my questions.

References

1. Johson, D., Perkins, C., Arkko, J.: Mobility Support in IPv6. RFC 3775 (June 2004)
2. Perkins, C.: IP Mobility Support for IPv4. RFC 3220 (January 2002)
3. Narten, T., Nordmark, E., Simpson, W.: Neighbor Discovery for IP Version 6 (IPv6). RFC 2461 (December 1998)
4. Dummore, M.: Final Mobile IPv6 Support Guide. 6NET IST-2001-32603 (December 2004)
5. Deering, S., Hinden, R.: Internet Protocol Version 6 (IPv6) Specification. RFC 2460 (December 1998)
6. Johnson, D., Deering, S.: Reserved IPv6 Subnet Anycast Addresses. RFC 2526 (March 1999)
7. Arkko, J., Kempf, J., Zill, B., Nikander, P.: SEcure Neighbor Discovery (SEND). IETF, RFC 3971 (2005)

RPKI Deployment: Risks and Alternative Solutions

Xiaowei Liu, Zhiwei Yan, Guanggang Geng, Xiaodong Lee,
Shian-Shyong Tseng and Ching-Heng Ku

Abstract In order to cope with the BGP (Border Gateway Protocol) security defects, RPKI (Resource Public Key Infrastructure) was proposed in IETF (Internet Engineering Task Force) in order to authenticate the relationship between IP prefix and its origination. Since 2012, a series of RPKI-related protocols have been standardized in IETF and the community has launched its actual deployment. However, with the global deployment of RPKI, a lot of concerns from technical, economic and political aspects have been raised. In this paper, we attempt to collect and analyze the most critical risks appeared during the RPKI deployment, and summarize the alternative solutions which have been presented to address or mitigate these risks.

1 Introduction

1.1 RPKI Architecture

On the Internet, BGP (Border Gateway Protocol) [1] is used for exchanging routing information between AS (Autonomous Systems)[1]. It also plays an

---

X. Liu(✉)· Z. Yan · G. Geng · X. Lee
China Internet Network Information Center, Beijing, China
e-mail: {liuxiaowei,yanzhiwei,gengguanggang,xl}@cnnic.cn

X. Liu· Z. Yan · G. Geng · X. Lee
National Engineering Laboratory for Naming and Addressing, Beijing, China

S.-S. Tseng
Asia University, Taichung, Taiwan
e-mail: sstseng@asia.edu.tw

C.-H. Ku
Taiwan Network Information Center, Taipei, Taiwan
e-mail: chku@twnic.net.tw

[1] An AS is a collection of IP prefixes which are controlled by one or more network operators. AS is the unit of inter-domain routing policy, and behaves as a single administrative entity (such as a university, an enterprise or a business division).

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_30

important part in making core routing decisions based on paths, network policies, or rules configured by network administrators.

However, BGP is vulnerable to a variety of routing attacks because of the lack of verification mechanisms to ensure the legitimacy of BGP signaling messages. In particular, the origin information of specific IP prefixes in a BGP message is fully exposed to accidental misconfigurations and deliberate attacks. The most common and devastating attack among them is prefix hijacking which is also referred to as BGP hijacking, or route hijacking [2]. Prefix hijacking is a misbehavior that an AS originates an IP prefix it does not own [3]. Over the past decade, the prefix hijacking incidents have occurred many times, such as in February 2008 Pakistan blocked YouTube traffic, in April 2010 China Telecom hijacked more than 37,000 IP prefixes and in February 2014 the Canadian ISP redirected traffic from nineteen ISPs. Prefix hijacking is becoming a serious security problem which can lead to black-hole routes, traffic interception and denial-of-service attacks on the global Internet [4]. RPKI (Resource Public Key Infrastructure) was mainly designed to cope with prefix hijacking attacks.

In RPKI, CAs (Certification Authorities) are organized in a hierarchical structure which is aligned to the existing INR (Internet Number Resources, including IP prefixes and AS numbers) allocation hierarchy [5]. Each INR allocation requires corresponding resource certificates to attest to it. Two important resource certificates are needed during this allocation process, and they are CA certificates and EE (End-entity) certificates: CA certificates attest to the INR holdings. EE certificates are primarily used for the validation of ROAs (Route Origin Authorizations) [6] which is used to verify whether an AS is permitted to originate routes to specific IP prefixes and Manifest [7] which is used to ensure the integrity of the repository [8].

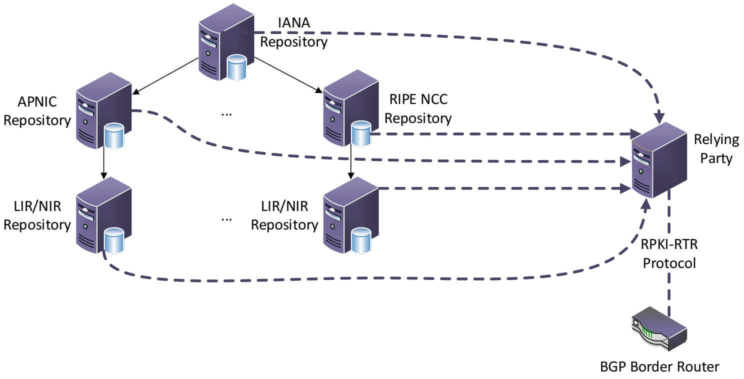


Fig. 1 Verify the route origin using RPKI

The process of using RPKI to verify the route origin is shown in Fig. 1.

- CAs, including IANA (Internet Assigned Numbers Authority), five RIRs (Regional Internet Registries), NIRs (National Internet Registries) and ISPs (Internet Service Providers), publish authoritative objects (including resource certificates, ROAs, Manifest and so on) into the their repositories.
- RPs (Relying Parties)[2] all over the world collect (with the rsync protocol [9]) and verify (using a validation tool, such as rcynic [10]) the PRKI objects from the distributed repositories, and provide the results of verification to BGP border routers.
- Finally, BGP border routers can make use of these results to verify the route origin information in the BGP messages they have received.

The basic architecture of RPKI was standardized by IETF in February 2012. Since then a series of RFC (Request for Comments) about RPKI have been published, such as the protocol to deliver validated prefix origin data to routers [11], the algorithm agility procedure for the RPKI [12], and operational considerations about the origin validation [13]. In addition, many extensions and improvements about RPKI are now under discussion and standardization, including RRDP (RPKI Repository Delta Protocol) [14] which is used to replace the current data synchronization protocol between RPs and repositories, a method [15] which secures RPSL (Routing Policy Specification Language) objects with RPKI, and BGPsec [16] which is designed to protect the BGP path security.

1.2 Status of RPKI Deployment

The five RIRs have finished the deployment of RPKI, and are now offering RPKI services to their members. A number of countries (Ecuador, Japan, Bangladesh, etc.) have also started to deploy RPKI interiorly. And in order to further promote the deployment of RPKI, ICANN (Internet Corporation for Assigned Names and Numbers), the five RIRs and a lot of countries have been making continuous efforts to solve the existing problems and improve the corresponding technical standards.

However, RPKI is still in its early stages of global deployment. According to the data provided by RPKI Dashboard [17], the current routing table holds about 589328 IP prefixes in total, and the RPKI validation state has been determined for 37213 IP prefixes, which means that only 6.31% of the prefixes in the routing table can be validated. And Table 1 offers a more intuitive display of the RPKI "adoption rate" (the percentage of members deployed RPKI) in the five RIRs.

---

[2] RPs are entities that make use of signed products from RPKI and perform the actual validation of them.

Table 1 Adoption rate of RPKI in 5 RIRs

RIR	AFRINIC	APNIC	ARIN	LACNIC	RIPE NCC
Adoption Rate	0.84%	1.4%	0.82%	25.24%	9.9%

As we can see from Table 1, LACNIC has the highest adoption rate, which is about 25.24%. While the adoption rates in ARIN and AFRINIC are much lower, which are only 0.84% and 0.82% respectively.

RIPE NCC provides some statistics regarding the number of resource certificates and ROAs in each RIR [18]. From these statistics we find a good sign that the global deployment status of RPKI rises gradually, and with its further evolution, the global adoption rate of RPKI will get a faster growth.

2 Risks of RPKI Deployment

During the process of incremental deployment of RPKI, a lot of technical, economic and political risks have appeared and will appear. In this section, we attempt to collect and analyze the risks which are most critical among them.

2.1 Technical Risks

2.1.1 More than One TA

A TA (Trust Anchor) is an authoritative entity represented by a public key and its associated data [19]. The public key acts as an authority to verify digital signatures. And the associated data describes the types of information and actions for which the TA is authoritative. There are more than one TA in the RPKI architecture today, IANA and five RIRs are candidates to be default TAs [20].

With more than one TA, there are risks that two or more organizations may accidentally or maliciously issue certificates for the allocation of the same INRs. Then it may lead to inconsistent and conflicting assertion about to whom the specific INRs have been allocated. This kind of risk obviously may cause resource conflicts on the Internet [21].

2.1.2 Risks of CAs

2.1.2.1 Misconfigurations

Considering misconfiguration is inevitable and the significant impact it may cause, misconfiguration of CAs in RPKI is a potential risk in deployment [22].

The misconfiguration of CAs in RPKI will lead to serious consequences similar to those caused by malicious attacks (black-hole routes, traffic interception, and denial-of-service attacks). For example, misconfigurations of an ROA (such as adding a new ROA [23], whacking an existing ROA [24]) will cause all routes covered by this ROA to become invalid.

2.1.2.2 Unilateral Resource Revocation

In the RPKI architecture, there is a risk that CAs have the power to unilaterally and abusively revoke the INRs which have been allocated to their descendents, just by revoking the corresponding CA certificates [20].

Since revocation of certificates usually leads to taking the resource holder offline, the risk of unilateral resource revocation is rather serious. And mechanisms to detect and prevent these malicious resource revocations are strongly needed.

2.1.3 Global Inconsistency

The risk of global inconsistency may also be called "mirror world attacks". In mirror world attacks, a malicious CA presents one view of the RPKI repository to some RPs, and a different view to others.

Since a CA in RPKI can control everything in its own repository, there are possibilities that a malicious CA may perform these attacks easily. For example, a malicious CA presents the correct view of its repository to some RPs, but a forged view (e.g. the CA adds a specific ROA deliberately) to the others. When these deceived RPs offer their validation results to BGP border routers, the routers may abandon the legitimate routes which are considered to be invalid according to the validation results they have received.

2.1.4 Data Synchronization

It is required in [20] that all repositories must be accessible via rsync [25] protocol which is used for RPs to get the RPKI objects in the global distributed repositories. However, the rsync protocol is considered to be controversial with its following disadvantages:

(i) *Lack of standards and non-modular implementation [26]*
Although rsync is widely adopted in backup, restore, and file transfer, it has not been standardized by IETF. And the rsync implementation is non-modular, making it difficult to use its code.

(ii) *Not good enough in efficiency, scalability and security [27]*
Rsync is efficient when it's used between one client and one server. However, in RPKI a large number of clients may regularly connect to the repository server at the same time. Rsync is not efficient and scalable enough to deal with this concurrent case. Moreover, rsync itself provides little security (no content encryption and weak authentication especially in old versions) while transferring data.

(iii) *Underlying overhead caused by repository updates during active data transmissions [28]*
During data transmissions between RPs and the repository, a new update to the repository may cause data inconsistency between them. And in order to rectify this inconsistency, extra overhead costs (such as performing the synchronization once more) are required.

2.1.5 Risks of Staged and Incomplete Deployment

Since the global deployment of RPKI is an incremental and staged process [29], a lot of unexpected issues may appear during this process. Let's take an example to explain one of these issues. As shown in Fig. 2, incomplete deployment of RPKI may cause legitimate routes to be misclassified into invalid. In Fig. 2, we make the following assumptions:

(i) CNNIC, ISP1 and ISP2 have deployed the RPKI, but ISP3 have not yet.
(ii) CNNIC allocated IP prefix 218.241.104.0/22 to ISP1 and 218.241.108.0/22 to ISP2.
(iii) Three ROAs (ROA1, ROA2, ROA3) are issued respectively by CNNIC, ISP1 and ISP2.

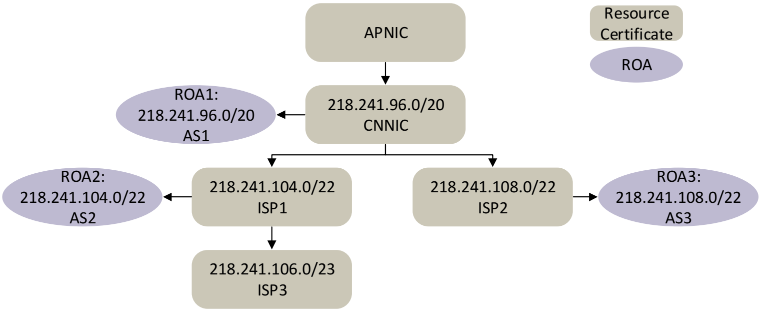


Fig. 2 An example of incomplete deployment

Now ISP3 announces to be the origination of 218.241.106.0/23. When other entities receive this announcement, they can validate it with ROAs information. Since prefix 218.241.104.0/22 described in ROA2 encompasses prefix 218.241.106.0/23 and no matching ROA describes 218.241.106.0/23 could be found [30], the announcement for prefix 218.241.106.0/23 will be considered to be invalid.

2.1.6 Low Validation Accuracy

The route origin validation accuracy refers to the percentage of valid routes. The calculation formula is shown in Equation (1).

$$Accuracy = \frac{number\ of\ valid\ routes}{number\ of\ valid\ routes + number\ of\ invalid\ routes} \tag{1}$$

As we can see from Table 2, the accuracy of route origin validation in the five RIRs differs a lot. LACNIC and RIPE NCC have the highest validation accuracy and both of them are over 90%, while the accuracy in AFRINIC is less than 50%. Many reasons may account for the low validation accuracy, such as misconfigurations, low adoption rates, etc.

Table 2 Route Origin Validation Accuracy in 5 RIRs

RIR	Total	Valid	Invalid	Unknown	Accuracy	Adoption Rate
AFRINIC	12795	52	56	12687	48.15%	0.84%
APNIC	143543	1399	667	141477	67.72%	1.44%
ARIN	207325	1362	335	205628	80.26%	0.82%
LACNIC	74840	17736	946	56158	94.94%	24.96%
RIPE NCC	147917	13324	1310	133283	91.05%	9.89%

When comparing the last two columns in Table 2, it can be found that a higher RPKI adoption rate tends to result in a higher validation accuracy. And this tendency can be explained by the example we present in section 2.1.5.

2.2 *Economic Risks*

To adopt RPKI, ISPs may need to deploy and maintain servers running as RPs, which means that ISPs have to take on more costs, including more responsibilities, more time and more money.

Besides, the route origin validation in RPKI has been designed to be deployed on current BGP border routers. In order to realize the validation, the border routers need to be upgraded to accommodate some additions, such as the "Route Validation Database" [31], which means that with the deployment of RPKI, there are additional costs to upgrade the network equipments.

2.3 *Political Risks*

RPKI provides a hierarchical architecture to improve the inter-domain routing security, however, it also induces power imbalance.

The most obvious example of power imbalance is Unilateral Resource Revocations which we have introduced in section 2.1.2.2. And this revocation may be driven by business disputes and political disagreements.

3 Alternative Solutions to RPKI Deployment Risks

In this section, we summarize and analyze the related alternative solutions and strategies to solve or mitigate the risks mentioned in Section 2.

3.1 Solutions to Technical Risks

3.1.1 Solutions to Multiple TAs

RIRs are trying to continually evolve RPKI, including the migration to a single GTA (Global Trust Anchor) as the root of the RPKI hierarchical structure. ICANN and RIRs have developed and deployed a technical testbed with an RPKI GTA [32]. It's assumed that there must be a single root trust anchor eventually [13]. With this single root trust anchor deployed, the risks of resource conflicts on the Internet could be reduced a lot.

However, this solution entitles more power to ICANN and exacerbates the risk of "power imbalance" mentioned in section 2.3.

3.1.2 Solutions to Misbehaved CAs

3.1.2.1 "Consent" Solution

To reduce the risk of unilateral resource revocation, Heilman et al. proposed a mechanism to improve the transparency of RPKI. They also offer some tools which can detect changes to the RPKI, for example, the CA revokes a resource certificate. The main idea of this mechanism is that any revocation of INRs requires consent from all impacted parties. And the consent comes in the form of a ".dead" object which is constructed recursively to ensure that every entities impacted could be notified [24].

This mechanism balances the power between the CAs in the RPKI hierarchical architecture to mitigate the risk of unilateral resource revocation.

3.1.2.2 "Suspenders" Solution

S. Kent et al. also put forward a collection of mechanisms named "Suspenders". "Suspenders" are designed to address the adverse effects on INR holders which were caused by CAs' accidental or deliberate misbehavior. This mechanism imports two new objects: an INRD (Internet Number Resource Declaration) file and a LOCK object. The INRD file is external to the RPKI repository, and it contains the most recent changes that were made by the INR holder. The LOCK object is published in the INR holder's repository. It contains a URL which points to the INRD file, and a public key used to verify the signature of INRD file. Whenever the RPs detect the inconsistencies between the actual changes and the INRD file, they can determine individually whether to accept these changes or not.

Both "Consent" solution and "Suspenders" solution are considered to be effective to protect RPKI against misconfigurations and malicious actions by authorities who abuse their power. However, "Suspenders" solution requires that INRD file be external to the RPKI repository. It may put more burdens on RPs to collect and verify these INRD files.

3.1.3 Solutions to Global Inconsistency

The mechanism provided by Heilman et al. in [24] is also effective to protect RPs against mirror world attacks. It allows RPs to raise an alarm whenever they found their views of RPKI repositories inconsistent with the others'.

Hristo Stoyanov [33] also provides a mechanism to reduce the global inconsistency risk. This mechanism constructs a k-connected spanning graph, and performs 2-party audits for all edges in the graph. This mechanism manages to ensure that every network device shares consistent information about the ownership of IP addresses to avoid mirror world attacks.

3.1.4 Solutions to Data Synchronization

A number of alternative protocols have been presented to take the place of "rsync" protocol due to its shortcomings mentioned above.

3.1.4.1 RRDP

T. Bruijnzeels et al. have proposed an alternative protocol (RRDP, RPKI Repository Delta Protocol) for RPs to keep their local caches in sync with the repository system [34]. This new protocol is based on notification, snapshot and delta files. When RPs query a repository for updates, they will use delta files (and snapshot files as needed) to keep their local caches updated. Moreover, RRDP protocol can work with the existing rsync URIs.

Compared with rsync protocol, RRDP is considered to be effective to eliminate a number of consistency related issues, help to reduce the load on publication servers, and have higher scalability.

3.1.4.2 Improved Rsync Protocol

CNNIC also proposed an improved rsync mechanism which transfers the work of checksums calculation to RPs in order to reduce the computation load on the rsync server side [35]. They also presented a NOTIFY mechanism that send NOTIFY message to key RPs to make them fetch the updated RPKI objects in time.

3.1.5 Solutions to Incomplete Deployment and Low Validation Accuracy

Both of the two risks (incomplete deployment and low validation accuracy) are caused by the partial deployment of RPKI. With the widely deployment of RPKI in the near future, these two risks ought to be mitigated.

3.2 Strategy to Reduce Economic Risks

Since what ISPs care about most is profit instead of security, it is essential to make ISPs aware of the economic interests behind the deployment. RPKI is the base of BGPsec [16] which can affect the route selection [36], and routers in a more secure AS which deployed with RPKI and BGPsec deserve a higher priority. With this strategy, an ISP who has deployed RPKI and BGPsec can attract more traffic, and then more revenue.

In terms of RPKI-capable border routers, a lot of network equipment manufacturers including Cisco, Juniper and Alcatel Lucent have put RPKI as an important feature in their new products [37]. And this fact could motivate more network equipment manufacturers to provide support for RPKI in order to keep pace with these pioneers.

3.3 Solutions to Mitigate Political Risks

Obviously, the "power imbalance" is both a technical and political risk. As mentioned in section 3.1.2, "Consent" mechanism and "Suspenders" mechanism are presented to address the unilateral resource revocation risk. And both of the two mechanisms have taken the political risks (such as government mandate and law enforcement actions) into consideration. So we think that the two mechanisms would be effective to mitigate the power imbalance risk.

4 Conclusion

Route origin validation offered by RPKI contributes much to the inter-domain routing security, and the corresponding technical standards has been improved a lot over time. During the process of global deployment of RPKI, a number of new problems and risks not restricted to technical (economic, political, etc.) have appeared. Although alternative solutions have been proposed continually, more efforts need to be done when considering that the global RPKI adoption rate is still very low. So far, there is no mechanism that could address all these risks mentioned. How to combine and choose the solutions analyzed in this paper is another important and significant challenge.

Besides, RPKI is only the first step to protect the inter-domain routing security, the next step is BGPsec. We will continue our work to collect and present the potential risks of BGPsec to attract more effective solutions.

Acknowledgments This work was supported by the Knowledge Innovation Program of the Chinese Academy of Sciences under Grant No. CNIC_QN_1405.

References

1. Rekhter, Y., Li, T., Hares, S.: A Border Gateway Protocol 4 (BGP-4). IETF RFC4271 (January 2006)
2. IP hijacking. http://en.wikipedia.org/wiki/IP_hijacking
3. Zhang, Z., Zhang, Y., Hu, Y.C., Mao, Z.M.: Practical Defenses Against BGP Prefix Hijacking. ACM CoNext (December 2007)

4. Ballani, H., Francis, P., Zhang, X.: A Study of Prefix Hijacking and Interception in the Internet. ACM SIGCOMM (2007)
5. Huston, G., Michaelson, G.: Validation of Route Origination Using the Resource Certificate Public Key Infrastructure (PKI) and Route Origin Authorizations (ROAs). IETF RFC 6483 (February 2012)
6. Lepinski, M., Kent, S., Kong, D.: A Profile for Route Origin Authorizations (ROAs). IETF RFC 6482 (February 2012)
7. Austein, R., Huston, G., Kent, S., Lepinski, M.: Manifests for the Resource Public Key Infrastructure (RPKI). IETF RFC 6486 (February 2012)
8. Huston, G., Loomans, R., Michaelson, G.: A Profile for Resource Certficate Repository Structure. IETF RFC 6481 (February 2012)
9. Weiler, S., Ward, D., Housley, R.: The rsync URI Scheme. IETF RFC 5781 (February 2010)
10. Rcynic. http://trac.rpki.net/wiki/doc/RPKI/RP
11. Bush, R., Austein, R.: The Resource Public Key Infrastructure (RPKI) to Router Protocol. IETF RFC6810 (January 2013)
12. Gagliano, R., Kent, S., Turner, S.: Algorithm Agility Procedure for the Resource Public Key Infrastructure (RPKI). IETF RFC6916 (April 2013)
13. Bush, R.: Origin Validation Operation Based on the Resource Public Key Infrastructure (RPKI). IETF RFC7115 (January 2014)
14. Bruijnzeels, T., Muravskiy, O., Weber, B., Austein, R., Mandelberg, D.: RPKI Repository Delta Protocol. draft-ietf-sidr-delta-protocol-00 (February 2015)
15. Kisteleki, R., Haberman, B.: Securing RPSL Objects with RPKI Signatures. draft-ietf-sidr-rpsl-sig-06.txt (November 2014)
16. Lepinski, M. (ed.): BGPsec Protocol Specification. draft-ietf-sidr-bgpsec-protocol-11 (January 2015)
17. RPKI Dashboard. http://rpki.surfnet.nl/global.html
18. RIPE NCC. http://certification-stats.ripe.net/
19. Housley, R., Ashmore, S., Wallace, C.: Trust Anchor Format. IETF RFC5914 (June 2010)
20. Lepinski, M., Kent, S.: An Infrastructure to Support Secure Internet Routing. IETF RFC 6480 (February 2012)
21. IAB statement on the RPKI. https://www.ietf.org/mail-archive/web/ietf-announce/current/msg07028.html
22. Malhotra, A., Goldberg, S.: RPKI vs ROVER Comparing the Risks of BGP Security Solutions. ACM SIGCOMM (2014)
23. Cooper, D., Heilman, E., Brogley, K., Reyzin, L., Goldberg, S.: On the Risk of Misbehaving RPKI Authorities. ACM Hotnets (November 2013)
24. Heilman, E., Cooper, D., Reyzin, L., Goldberg, S.: From the Consent of the Routed-Improving the Transparency of the RPKI. ACM SIGCOMM (2014)
25. rsync web pages. https://rsync.samba.org/
26. Oleg Muravskiy: RPKI Repository Analysis and Delta Protocol. http://www.ietf.org/proceedings/86/slides/slides-86-sidr-2.pdf
27. rsync considered inefficient and harmful. https://www.ietf.org/proceedings/89/slides/slides-89-sidr-6.pdf
28. Weber, B.: RPKI Repository Distribution Protocol(RRDP). https://www.ietf.org/mail-archive/web/sidr/current/msg05367.html
29. George, W.: Adventures in RPKI (non)deployment. https://www.nanog.org/sites/default/files/wednesday_george_adventuresinrpki_62.9.pdf

30. Huston, G., Michaelson, G., Loomans, R.: A Profile for X.509 PKIX Resource Certificates. IETF RFC6487 (February 2012)
31. Example: Configuring Origin Validation for BGP. https://www.juniper.net/documentation/en_US/junos12.2/topics/topic-map/bgp-origin-as-validation.html
32. GTA testbed. https://myicann.org/plan/project/5283e47c0038d63c92a626c2f26a59f6
33. Stoyanov, H.: Cryptographically secure detection of mirror worlds. http://web.mit.edu/rsi/2014/all/hristo.pdf.gz
34. Bruijnzeels, T., Muravskiy, O., Weber, B.: RPKI Repository Analysis and Requirements. draft-tbruijnzeels-sidr-repo-analysis-00 (February 2013)
35. Wang, C., Yan, Z., Hu, A.: An Efficient Data Management Architecture for the Large-scale Deployment of Resource Public Key Infrastructure. IEEE CECNet (December 2014)
36. Gill, P., Schapira, M., Goldberg, S.: Let the Market Drive Deployment: A Strategy for Transitioning to BGP Security. ACM SIGCOMM (2011)
37. Resource Public Key Infrastructure (RPKI). http://www.slideshare.net/SienaPerry/introduction-to-rpki-rpki-my-nog20140821shortv2

Building the Virtual Experiment Learning Activities to Facilitate IPv6 Online Training

Jun-Ming Su, Shian-Shyong Tseng and Ting-Wei Lin

Abstract Due to the IPV4 address exhaustion problem, how to efficiently promote and learn the IPv6 upgrade knowledge has been paid more attention. Therefore, based on the scaffolding concept, this study proposes a design scheme of the Virtual Experiment Learning Activity for IPv6 (IPv6-VELA) to integrate learning materials for conceptual learning with virtual experiment for hands-on learning. Thereby, the trainee can learn the IPv6 upgrade knowledge in the step-by-step manner by means of the scaffolding learning supports, i.e., formative and summative diagnostic reports generated according to the learning portfolio. The experimental result also shows that the IPv6-VELA-based learning activity is beneficial for the trainee and teacher.

Keywords Ipv6 upgrade · Virtual experiments · Online training · Learning activity · Scaffolding · Personalized diagnosis

1 Introduction

The rapid development of Internet applications has led to the problem of IPV4 address exhaustion [1,4]. Therefore, how to efficiently learn the IPv6 upgrade knowledge has become an important issue [2]. During the training activity of the IPv6 upgrade, trainee is required not only to learn the Substantive Knowledge, e.g., the physical meaning of IPv6 route, but also to practice the Hands-on Procedural Knowledge, e.g., the setting process of IPv6 route [5,6,7]. However, most of

---

J.-M. Su(✉) · T.-W. Lin
Department of Information and Learning Technology,
National University of Tainan, Tainan, Taiwan
e-mail: junming.su@gmail.com, ads079921@hotmail.com

S.-S. Tseng
Department of Applied Informatics and Multimedia, Asia University, Taichung, Taiwan
e-mail: sstseng@asia.edu.tw

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_31

IPv6 learning content is able to describe the operational process of IPv6 upgrade, but the actual hand-on operation is not able to be efficiently supported. Thus, the simulation environments have been proven to be effective for the hand-on operation learning [9,10]. Nevertheless, the much higher fidelity the simulation environment is able to offer, the much higher learning skills the trainee is required to have. Consequently, the trainee is not easy to learn the Hands-on Procedural Knowledge by means of the higher fidelity simulation environments due to the higher required learning skills, which thus leads to the learning barriers. Moreover, the most of existing simulation environments lack the diagnostic report generated automatically according to the operation portfolio, so the trainees and teachers are difficult to understand the learning problems for the self-learning and in-class instruction, respectively.

Therefore, this paper aims to propose a design scheme of the Virtual Experiment Learning Activity for IPv6 (IPv6-VELA) to offer trainees the progressive scaffolding learning strategy [11]. The IPv6-VELA scheme is able to design a learning activity to integrate the learning content with Diagnostic Virtual Experiment (DiVE) [14]. Thereby, trainees can learn the conceptual knowledge by the learning content and virtually practice the hands-on procedural knowledge by the DiVE. Furthermore, during the learning activity, they can read the Formative Diagnostic Report (FDR) [12], the personalized diagnostic report that DiVE automatically generated, to instantly understand the operational problems and to further improve them. After the finish of the learning activity, the Summative Diagnostic Report (SDR) [13], the personalized diagnostic report generated automatically according to learning activity's operational portfolio, to understand the learning status and problems in-depth for the self-reflection and learning improvement.

2 Related Works

Internet Protocol version 4 (IPv4), a 32-bit addresses protocol described in IETF publication RFC 791 [1], is defined to communicate and route the most traffic on the Internet over 20 years old. Accordingly, Internet Protocol version 6 (IPv6) [2], a 128-bit addresses protocol developed by the Internet Engineering Task Force (IETF) [3], is defined solve IPv4 address exhaustion problem, i.e., the shortage of the IPv6 address [4]. Therefore, how to train relevant persons in charge of the network management to promote the IPv6 specification in the early stage has indeed become an important issue. Many IPv6 books [17] and online learning resources [15,18] have thus been published and announced for the IPv6 knowledge learning. However, most of the IPv6 knowledge can be regarded as the procedural knowledge, which are required to perform the hands-on operation for mastering the learning and not only the substantive/conceptual knowledge. The simulation tools, which allow users to virtually conduct the virtual experiments [9,10] to experience the hands-on learning, have shown the effectiveness on the procedural knowledge learning [5,6,7]. Therefore, several simulation tools have also been developed to assist users in learning IPv6 upgrade. For example, the

Cisco Packet Tracer [16], current version is 6.1, can be used to create networks for the Router Simulation in terms of IPv6. Thus, Danscourses [15] uses the Cisco Packet Tracer to design several IPv6-related learning content to assist users in self-leaning the complicated technological concepts. However, it may not be suitable for the beginners and novice to self-learn the IPv6 upgrade knowledge in a step by step manner because its operations usually involve many prerequisite knowledge and it lacks the automatic diagnosis to analyze the learning problems. In general, the much higher fidelity the simulation environment is able to offer, the much higher learning skills the trainee is required to have. Accordingly, how to design the suitable learning content to facilitate the IPv6 upgrade online training is our main concern in this paper.

3 The Design Scheme of Virtual Experiment Learning Activities for IPv6 (IPv6-VELA)

3.1 The Scaffolding Strategy of IPv6-VELA

In order to offer the efficient training method of IPv6 upgrade, this paper proposed the design scheme of the Virtual Experiment Learning Activity for IPv6 (IPv6-VELA) to perform the Progressive Scaffolding learning strategy. As shown the definition of IPv6-VELA, the design scheme integrates Course Unit (CU) with Diagnostic Virtual Experiment (DiVE) [14], where trainee can learn the Conceptual/Substantive Knowledge by the CUs, learn the Hands-on Procedural Knowledge by the DiVEs, and understand and reflect the learning problems by the diagnostic report of DiVE. Accordingly, the (self-)learning performance can thus be improved through the progressive scaffolding learning supports.

The Definition of IPv6-VELA:

- IPv6-VELA = {CU_{set}, $DiVE_{set}$, FDR, SDR}, where
  (a) CU_{set} = (CU_1, CU_2,…, CU_n): a set consists of n course units (CU_n), each of which can be the learning materials or the test item and has associated Conceptual Knowledge (CK) related to the IPv6 upgrade subject.
  (b) $DiVE_{set}$ = ($DiVE_1$, $DiVE_2$,…, $DiVE_m$): a set consists of m DiVEs, each of which has associated Conceptual Knowledge (CK) and Procedural Knowledge (PK) and can provide the online virtual operations and diagnostic report,
  (c) FDR: denotes the Formative Diagnostic Report (FDR) generated automatically according the operational portfolio of a DiVE,
  (d) SDR: denotes the Summative Diagnostic Report (SDR) generated automatically according the learning portfolio of an IPv6-VELA.

Fig. 1 illustrates the definition and example of the IPv6-VELA scheme. Therefore, an IPv6-VELA is able to be designed to comprise n number of CUs and m number of DiVEs to offer the learning of Conceptual Knowledge (CK) and operational practice of Procedural Knowledge (PK) according to training requirements, respectively. Furthermore, FDR, the personalized diagnostic report of a finished DiVE, can provide trainee with the formative suggestion [12] to efficiently understand the operational problems. Consequently, trainee can self-decide to re-read the relevant CUs and re-try the DiVE, or continue the next advanced learning steps, whereby they can master the required CK and PK by the progressive scaffolding learning strategy [11]. Finally, SDR [13], the personalized diagnostic report of a finished IPv6-VELA, is able to make trainee to understand the learning barriers and assist them for the self-reflection and learning improvement.

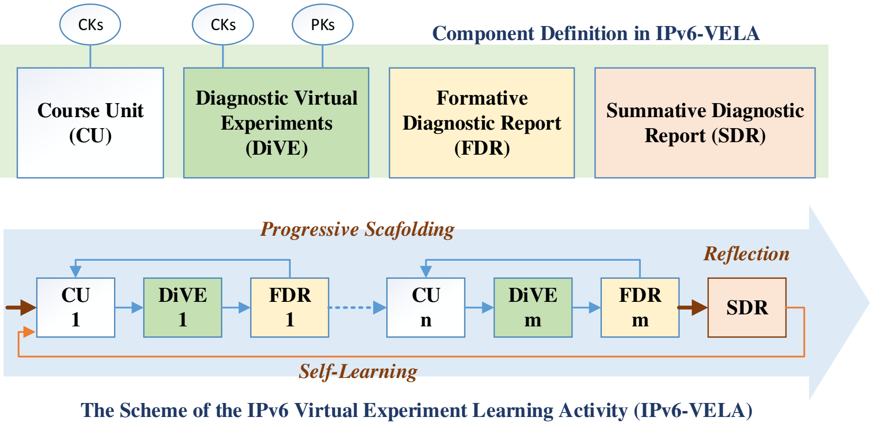

CKs
CKs
PKs
Component Definition in IPv6-VELA
Course Unit (CU)
Diagnostic Virtual Experiments (DiVE)
Formative Diagnostic Report (FDR)
Summative Diagnostic Report (SDR)
Progressive Scaffolding
Reflection
CU 1
DiVE 1
FDR 1
CU n
DiVE m
FDR m
SDR
Self-Learning
The Scheme of the IPv6 Virtual Experiment Learning Activity (IPv6-VELA)

Fig. 1 The Definition and Example of the IPv6-VELA

3.2 The Learning Procedure of IPV6-VELA

Based on the proposed IPv6-VELA scheme, an IPv6-VELA-based learning activity can thus be created to integrated required course unit (CUs) with DiVEs, which is able to be associated with relevant CKs and PKs for the diagnosis. Fig. 2 shows an example of IPv6-VELA, where the left-hand side of the Fig. 2.a shows the structure of the learning activity and the right-hand side of the Fig. 2.a shows the Learning Materials for the Conceptual Knowledge (CK) learning, e.g., CU_1. The Fig. 2.b illustrates the operational process of a DiVE for the Procedural Knowledge (PK) learning, whereby trainee is able to virtually operate the IPv6 setting process using the Web-based interface.

After the operation of a DiVE, a personalized diagnostic report, i.e., FDR, will be generated automatically according to the analysis of the DiVE's operation records. As shown in Fig. 3.a, the total score of the operated DiVE is shown on the top, and the trainee's online recorded operation video with the teacher's corresponding correct operation video are also shown for the comparison and self-reflection. Moreover, the action sequences, each of which consists of key actions, that are required to be operated in a DiVE are listed to easily and visually understand the operation procedure of a DiVE for trainees. A key action will be marked with green color if the trainee did the right operation, otherwise, it will be marked with red color. For example, the action 2 with red color is an error operation and its error action description will be shown if trainee click on it for understanding the problems. After the finish of a learning activity, a SDR will be generated as well to assist trainees in understanding the learning problems and recall the portfolio concerning the CUs and DiVEs of an IPv6-VELA-based learning activity for the learning improvement.

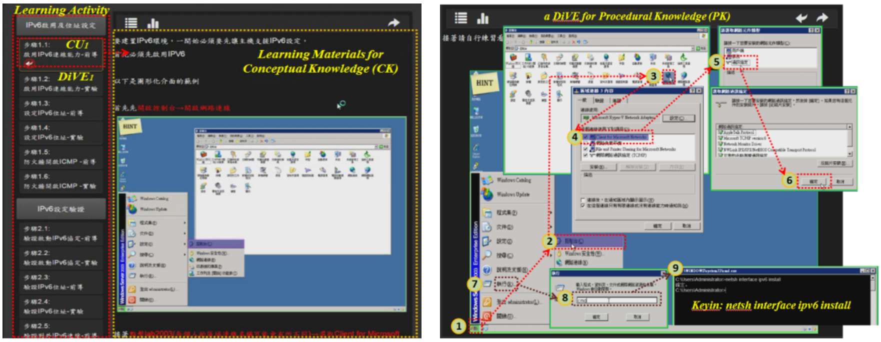

Learning Activity
CU1
DiVE1
Learning Materials for Conceptual Knowledge (CK)
HINT
a DiVE for Procedural Knowledge (PK)
1
2
3
4
5
6
7
8
9
Keyin: netsh interface ipv6 install

Fig. 2 (a) a IPv6-VELA-based Learning Activity, (b) a DiVE for hands-on learning

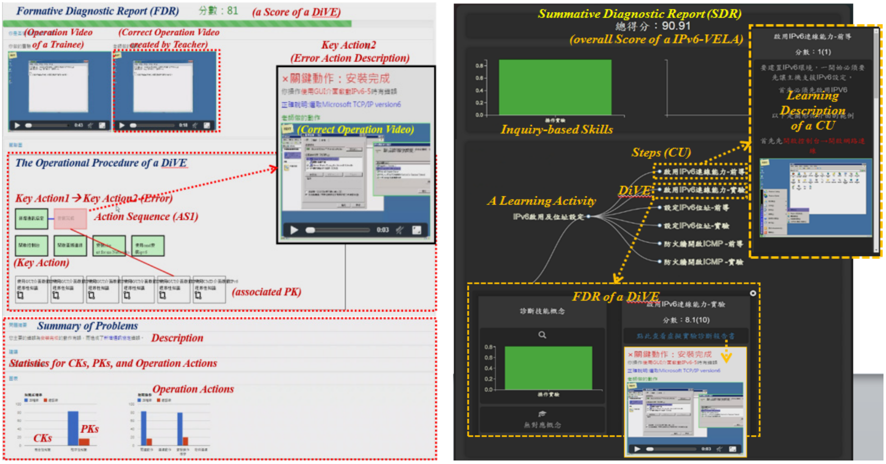

Formative Diagnostic Report (FDR)
分數：81
(a Score of a DiVE)
(Operation Video of a Trainee)
(Correct Operation Video created by Teacher)
Key Action2
(Error Action Description)
×關鍵動作：安裝完成
(Correct Operation Video)
The Operational Procedure of a DiVE
Key Action1 → Key Action2 (Error)
Action Sequence (AS1)
(Key Action)
(associated PK)
Summary of Problems
Description
Statistics for CKs, PKs, and Operation Actions
Operation Actions
CKs
PKs
Summative Diagnostic Report (SDR)
總得分：90.91
(overall Score of a IPv6-VELA)
Inquiry-based Skills
Steps (CU)
A Learning Activity
DiVE
分數：1(1)
Learning Description of a CU
FDR of a DiVE
分數：8.1(10)

Fig. 3 (a) Formative Diagnostic Report (FDR), (b) Summative Diagnostic Report (SDR)

4 Experiment and Analysis

4.1 Experimental Design

In order to evaluate the effectiveness of the proposed IPv6-VELA scheme, the experiment was conducted. Twenty-nine trainee participated the IPv6 training course taught by a teacher for six hours in one day. The learning subject of the experimental learning activity based on IPv6-VELA is “IPv6 Route Setup”, which consists of four sub-activities with several steps, i.e., (1) Basic Setup, (2) Static Route Setup, (3) RIP Dynamic Route, and (4) OSPF Dynamic Route. This IPv6-VELA-based learning activity was used by teacher for the instruction and by trainee for the self-practice and self-learning as shown in Fig. 4.

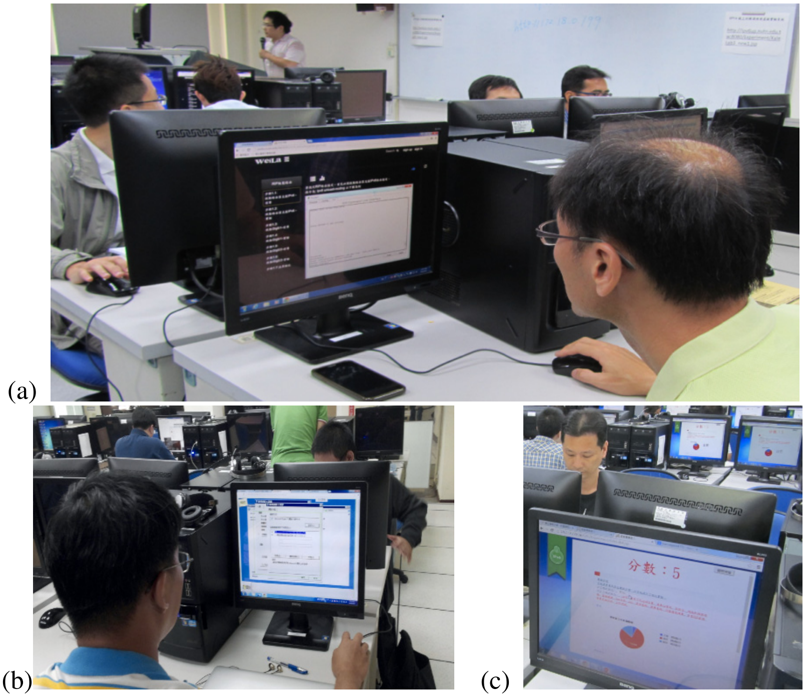


Fig. 4 (a) Teacher teaching and Trainee reading the course unit (CU), (b) Operating the DiVE, and (c) Reading the Diagnostic Report

4.2 Questionnaire Analysis

After conducting the experimental IPv6 training course, the questionnaires for trainee, consisting of seven aspects, i.e., (1) Learning Satisfaction, (2) Learning Motivation, (3) Usefulness of Diagnostic Report, (4) Willingness for future use, (5) Perceived usefulness, (6) Perceived ease of use, and (7) Cognitive Load, had also been defined to collect the degrees of learning feedbacks. As shown in Fig. 5, the statistic results of the feedback analysis indicate that trainee’s overall degree of learning satisfaction gains a higher satisfaction score of 4.24 (a score of 3.0 is neutral) and they had high learning motivation to use the IPv6-VELA-based learning activity (a score of 3.97) and were highly satisfied with the learning support due to the Usefulness of Diagnostic Report (a score of 4.15) and Perceived usefulness (a score of 4.03). Accordingly, they were highly willing to the future use and recommend the IPv6-VELA-based learning activity to others (a score of 4.33).

Moreover, the IPv6-VELA-based learning activity was easy to use due to the high score of 4.03 for the Perceived ease of use aspect and the low score of 29.08 (100 is highest) for the Cognitive Load aspect.

In addition, the teacher also agreed that the IPv6-VELA-based learning activity can efficiently assist trainee in understanding and self-reflecting the conceptual (substantive) and procedural knowledge in a step by step manner. Especially, some of the IPv6 simulation-based environments are not easy to be deployed and used for the classroom instruction, and the teaching performance can thus be improved due to the personalized diagnostic report generated automatically.

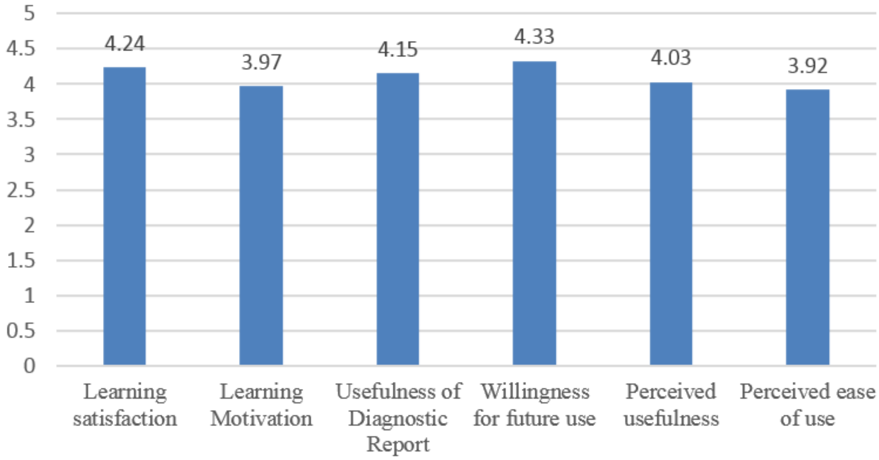


Fig. 5 Statistical results of the satisfaction questionnaire for trainee

5 Conclusion

This paper proposed a Virtual Experiment Learning Activity for IPv6 (IPv6-VELA) scheme to create a learning activity that is able to integrate learning materials with Diagnostic Virtual Experiment (DiVE). Thereby, the trainee can read the learning material and then virtually operate the virtual experiment. Furthermore, the learning effectiveness can be progressively scaffolded by the FDR and SDR supports. According to the experiments, the trainees' feedback show the higher satisfaction score in terms of the (1) Learning Satisfaction (a score of 4.24), (2) Learning Motivation (a score of 3.97), (3) Usefulness of Diagnostic Report (a score of 4.15), (4) Willingness for future use (a score of 4.33), (5) Perceived usefulness (a score of 4.03), (6) Perceived ease of use (a score of 4.03), and (7) Cognitive Load (a score of 29.08). In the near future, the flexibility of the IPv6-VELA's structure, the power of the diagnosis concerning the FDR and SDR are going to be improved.

Acknowledgement This paper is partially sponsored by National Science Council of Republic of China under the number of NSC NSC 102-2511-S-468-003-MY2 and NSC 102-2511-S-024-009-MY2, Ministry of Transportation and Communications, Republic of China, under Grant 1040212coco, and Taiwan Network Information Center (TWNIC).

References

1. RFC 791, Internet Protocol, DARPA Internet Program Protocol Specification, Defense Advanced Research Projects Agency (Stepember 1981)
2. Deering, S., Hinden, R.: RFC 2460, Internet Protocol, Version 6 (IPv6) Specification (December 1998)
3. IETF, Internet Engineering Task Force (retrieved 2015). https://www.ripe.net/participate/internet-governance/internet-technical-community/ietf
4. IPv4 address exhaustion (retrieved 2015). https://en.wikipedia.org/wiki/IPv4_address_exhaustion
5. Bennett, R.E., Persky, H., Weiss, A., Jenkins, F.: Measuring problem solving with technology: A demonstration study for NAEP. The Journal of Technology, Learning, and Assessment **8**(8), 1–44 (2010) (retrieved from). http://www.jtla.org
6. Jacobs-Sera, D., Hatfull, G.F., Hanauer, D.I.: Active Assessment: Assessing Scientific Inquiry. Mentoring in Academia and Industry **2**(2009), 31–43 (2009)
7. Wenning, C.: Assessing inquiry skills as a component of scientific literacy. Journal of Physics Education Online **4**(2), 21–24 (2007)
8. Molecular WorkBench (2015). http://mw.concord.org/modeler/
9. Finkelstein, N.D., Adams, W.K., Keller, C.J., Kohl, P.B., Perkins, K.K., Podolefsky, N.S., et al.: When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. Physics Review Special Topics–Physics Education Research **1**, 010103-1–010103-8 (2005)
10. Zacharia, Z.C., Olympiou, G., Papaevripidou, M.: Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. Journal of Research in Science Teaching **45**(9), 1021–1035 (2008)
11. Wood, D., Bruner, J.S., Roos, G.: The Role of Tutoring in Problem Solving. J. Child Psychol. Psychiat. **17**, 89–100 (1976)
12. Formative assessment, (retrieved 2015). https://en.wikipedia.org/wiki/Formative_assessment
13. Summative assessment, (retrieved 2015). https://en.wikipedia.org/wiki/Summative_assessment
14. Su, J.M., Huang, J.R.: Building a Web-based Visualized Authoring System to Collaboratively Create Reconfigurable and Diagnostic Virtual Experiments in Natural Science. In: The 4th International Conference on Learning Technologies and Learning Environments (LTLE 2015) in Conjunction with 4th International Congress on Advanced Applied Informatics, July 12–16. Okayama Convention Center, Okayama, Japan (2015)
15. Danscourses, (retrieved 2015). http://www.danscourses.com/
16. Packet Tracer, "An innovative network configuration simulation tool free for Networking Academy students" (retrieved 2015). https://www.netacad.com/about-networking-academy/packet-tracer
17. Dooley, M., Rooney, T.:I Pv6 Deployment and Management, 1st edn. Wiley-IEEE Press (May 6, 2013)
18. IPv6 Deployment and Support, (retrieved 2015). http://www.6deploy.eu/index.php?page=e-learning2

Defense Denial-of Service Attacks on IPv6 Wireless Sensor Networks

Chia-Mei Chen, Sung-Chien Hsu and Gu-Hsin Lai

Abstract As IoT (Internet of Things) has become popular, more tiny devices and sensors are deployed to monitor our environment. Wireless sensor network is a typical implementation of IoT. Wireless sensor networks consist of many tiny sensors with sensing and communication ability. The applications of wireless sensor networks are widely used today in battlefield guard, circumstance monitor, and traffic analysis. Point to point communication is one of the most important characteristics of IoT, and a great volume of IP address is needed for IoT applications. The existing IP protocol, IPv4, could not provide enough public IP addresses for IoT applications. IPv6 allowing $2^{128,}$ or approximately 3.4×10^{38} addresses, or more than 7.9×10^{28} times as many as IPv4. It is a promising solution for IoT applications and IP shortage issues. Sensor network often carries on mission-critical tasks. Thus, security in sensor network is vital. Wormhole attack may threat the availability of networks by dropping data randomly or disturbing routing path. In this paper, we propose a detection mechanism which is based on RPL routing protocol. In RPL routing protocol, an information called “Rank” is used to avoid routing loops by computing a node’s position relative to other nodes with respect to the DODAG(Destination Oriented Directed Acyclic Graph) root. The proposed approach detects wormhole attack by identifying abnormal Rank values. The proposed approach uses only Rank information without any extra hardware.

Keywords Wormhole attack · Ipv6 · Wireless sensor network · RPL

C.-M. Chen · S.-C. Hsu
Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan
e-mail: cchen@mail.nsysu.edu.tw

G.-H. Lai(✉)
Department of Information Management, Chinese Culture University, Taipei, Taiwan
e-mail: Lgx4@ulive.pccu.edu.tw

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_32

1 Introduction

The rapid development of Internet of things (IoT) technology makes it possible for connecting a great volume of sensors together through the Internet. Wireless sensor network is a typical implementation of IoT. Wireless sensor network is suitable for several areas like agricultural areas, manufacturing industry or service industry. A sensor network contains a lot of nodes. Each node equips with a sensor to collect the data from the environment. Because sensor network often carries on mission-critical task, loss of availability may have serious impacts (e.g. fail to detect a potential accident and result in financial loss in plant monitor or environmental surveillance). To commutate with other sensors, each sensor need a public IP address. However, existing IP protocol, IPv4, could provide just about 2^{32} IP addresses. IP shortage problems become a critical issues for the development of IoT. IPv6 provides not only enough address space but also extra features like IPSec, QoS, End-to-end Connectivity and Auto-configuration. It seems IPv6 is the best solution for IP shortage problem and IoT, and some IoT standard like Zigbee [11] and Thread [9] already support IPv6 protocol. Zigbee is an open, global wireless standard to provide the foundation for the Internet of Things by enabling simple and smart objects to work together.

Although IPv6 provides a great number of IP addresses and extra features, adopting IPv6 could raise some security issues. Attackers could use the vulnerability of IPv6 protocol like ICMPv6 or extension header to damage IPv6 networks. Wormhole attack is common used to damage IPv6 network. In the wormhole attack, two malicious sensor nodes may be deployed and will construct a wormhole tunnel through a low-latency communication link, replaying messages from neighbors and dropping messages to disturb the normal function of the network. Victim sensors would think the routing path through the tunnel is shorter than the normal multi-hop route, as the wormhole tunnel reaches further than the normal wireless transmission range of a single hop [6]. In IPv6 network, it is very easy for attackers to launch wormhole attack. Malicious nodes could construct wormhole tunnel by sending out fake router advertisement. In this paper, we use the information provided by RPL routing protocol. IoT nodes in the proposed mechanism will ignore packets with unreasonable Rank sent by malicious nodes. The proposed system use existing routing protocol and no additional hardware like GPS or directional antenna is needed

2 Related Work

Wormhole attack in wireless ad-hoc networks was first independently introduced by Hu [4], Dahill [3], and Papadimitratos [8].

Hu et al. proposed packet leash mechanism against wormhole attacks [4-6]. Two different leashes: geographic leashes and temporal leashes were proposed to restrict the packet's maximum allowed transmission distance. However, to construct a geographical leash, in general, each node must know its own location, and all nodes must have loosely synchronized clocks, and all nodes must have tightly synchronized clocks to construct a temporal leash. Khalil et al. proposed a routing protocol called MobiWorp to detect and isolate wormhole tunnel. MobiWorp rely on a secure central authority (CA) for global tracking of node positions. MobiWorp also deployed a special node called guard node to maintain a black list and monitor network traffic [7]. Using GPS or other special hardware is not practical in IoT environment because of its expensive cost. Moreover, for some indoor application, nodes of sensor network could not receive GPS signal. Additional special nodes like guard node also need additional cost. Therefore, a detection mechanism without special node or hardware is needed for wireless sensor network.

More and more sensor network standard like Zigbee[11] or Thread [9] support IPv6. Therefore, most devices of sensor network are vulnerable from IPv6 attacks. Caicedo and Joshi [1] categorized IPv6 attacks into four types, they are reconnaissance attacks, host initialization and associated attack, attacks using routing headers and multicast-based attacks [1]. The details of these different attacks are summarized in Table 1.

Table 1 Summary of IPv6 attack from Caicedo and Joshi [1]

Attack Approach	Description
Reconnaissance attacks	Host probing and port scanning are usually the initial activities an attacker engages in to discover vulnerabilities in a network. In host probing, the attacker tries to identify the hosts connected to a network [a1]. Attackers could tools like nmap to launch reconnaissance attacks
DoS attack on DAD protocol	Victim first send a NS package to multicast address. The attacker falsely responds with a NA packet, informing the new node that it is already using that address. The new node generates another address and repeats the DAD procedure; the attacker again falsely responds with an NA packet. Eventually, the new node gives up without initializing its interface
Man-in-the-middle attack	When victim wants to transmit data to other node in LAN, victim first sends an NS message to the all-nodes multicast address. Attacker will reply to it with the corresponding NA message, thereby taking over the intended traffic flow between Node victim and other node. Attackers could steal confidential data.
Bogus router implantation attack	Attackers will propagate bogus address prefix information to reroute legitimate traffic to prevent the victim from accessing the desired network.
Attacks using routing headers	An attacker can generate specific packets with routing headers to reach hosts that normally would not accept the attacker's traffic
Multicast-based attacks	An attacker can launch a DoS attack on a multicast group by sending messages to the group address telling members to leave

Choudhary and Sekelsky [2] also listed some IPv6 vulnerabilities, they are:

- Hop-by-hop options.
- Auto-configuration
- Multiple Addresses
- ICMPv6 Filtering

According to prior works, most attacks use bogus or malicious ICMPv6 messages. In this paper, we will use ICMPv6 to detect wormhole attack in sensor network.

3 Proposed System

In this paper, we propose an intrusion detection system to detect and prevent wormlole and flooding attack in IPv6 sensor network environment. Figure 1 illustrates the framework and processes of the proposed system. The details of the processes are described as follows:

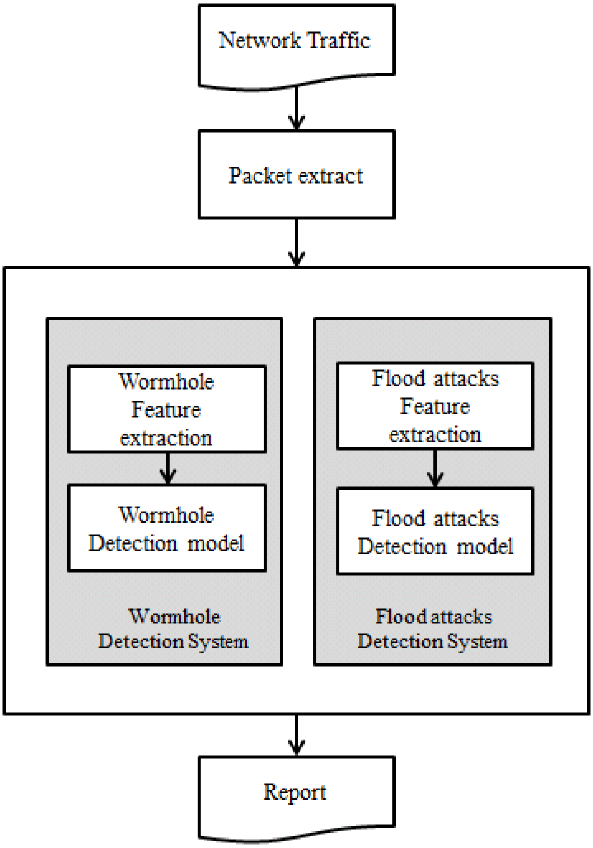

Network Traffic
Packet extract
Wormhole Feature extraction
Wormhole Detection model
Wormhole Detection System
Flood attacks Feature extraction
Flood attacks Detection model
Flood attacks Detection System
Report

Fig. 1 Framework and processes of proposed system

To detect and prevent wormhole attack, the proposed system is based on RPL Routing Protocol. In RPL, The RPL specification defines four types of control messages for topology maintenance and information exchange. The first one is called DODAG Information Object (DIO) and is the main source of routing control information. It may store information like the current Rank of a node, the current RPL Instance, the IPv6 address of the root, etc [10]. The proposed system firstly collects DIO packets to extract Rank message in DIO. Rank message is used to avoid routing loops. It means a node's position relative to other nodes with respect to the DODAG root. The proposed system extracts Rank message and monitor messages whether they contain abnormal Rank information. Figure 2 illustrates the process of the wormhole detection model.

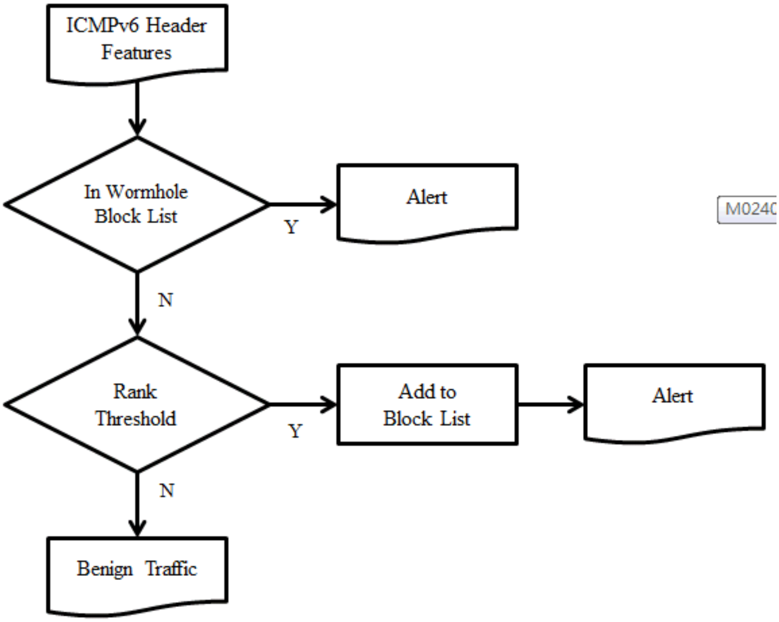


Fig. 2 Process of wormhole detection model.

Rather than analyzing all traffic with in network, the proposed system only collects ICMPv6 packets. While sensors receive ICMPv6 messages, sensors will check if the sender is a member of the wormhole black list. If the ICMPv6 message is from a node in black list, the receivers will issue an alert. Receivers will then check Rank value in ICMPv6 message. If Rank value in ICMPv6 message is unreasonable (too large), the sender is identified as attacker, the receivers will issue an alert. Figure 3 shows the pseudo code of the wormhole detection system.

To detect flooding attack, the proposed system focuses on abnormal flooding router advertisements and neighbor advertisements attacks. The proposed system will firstly check if the messages come from nodes in black list. Then The proposed system will count number of router advertisements and neighbor advertisements messages. When the number exceeds a threshold, the node will discard malicious router advertisements and neighbor advertisements messages. Figure 4 illustrates the process of flooding attack detection model.

```
threshold = USER DEFINED;      //
rank_diff = parent's rank – source rank;
if(target.isInWormholeTunnel()){
    if (rank_diff >= threshold) {
        true positive;
    }else{
        false negative;
    }
}else if(!target.isInWormholeTunnel()){
    if (rank_diff >= threshold) {
        false positive;
    }else{
        true negtive;
    }
}
```

Fig. 3 Pseudocode of wormhole detection system.

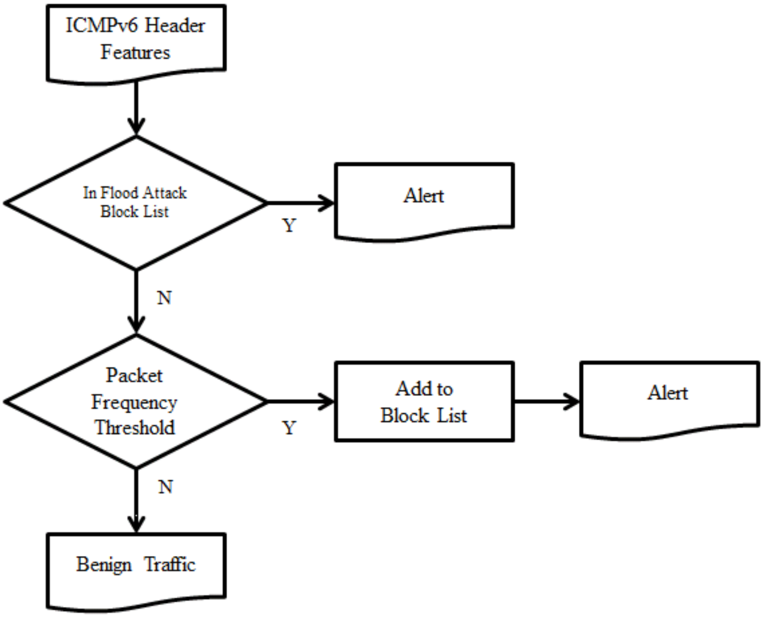

ICMPv6 Header Features
In Flood Attack Block List
Y
Alert
N
Packet Frequency Threshold
Y
Add to Block List
Alert
N
Benign Traffic

Fig. 4 process of detecting flooding attack.

4 Simulation and Result

In this section, we present the simulation environment and results for the proposed approach. The goal of this simulation is to evaluate the performance of the proposed method. For wormhole detection, our system try to detect malicious ICMPv6 messages correctly. Confusion matrix is used as measurements and in shown in table 1.

Table 2 Confusion matrix

	Identified as Malicious	Identified as Benign
Malicious ICMPv6	TP	FP
Benign ICMPv6	FN	TN

In this paper, we use four type of measurement to evaluate the performance of proposed approach, they are precision (SP), recall (SR), accuracy (A) and miss rate (MR). The formulas for measurements are as below:

$$SP = \frac{TP}{TP + FP} \tag{1}$$

$$SR = \frac{TP}{TP + FN} \tag{2}$$

$$A = \frac{TN + TP}{TP + FN + FP + TN} \tag{3}$$

$$MR = \frac{FP}{TN + FN} \tag{4}$$

Table 3 illustrates the parameters of simulation, and table 4 demonstrates the result of this simulation. The result of simulation shows that the proposed could detect wormhole correctly.

Table 3 Parameters of simulation

Map Size	500*500 (M)
Number of nodes	100
Transmission distance	100 M
Distance of wormhole tunnel	200 M

Table 4 Simulation result

SP	100%
SR	100%
A	100%
MR	0

5 Conclusion

Because IoT will be the trend, more and more tiny, smart devices will connect to each other. IPv6 is the next generation IP standard which provides almost infinite IP addresses. However, IPv6 also suffers from wormholes and flooding attacks. The proposed system is based on the RPL routing protocol, and only uses very simple information without additional hardware. The simulation shows the proposed system could detect wormholes correctly. The proposed system only focus on the availability of IPv6 based sensor networks. However, the confidentiality is also important for the application of IoT. The future work should design a complete mechanism which considers both availability and confidentiality together to secure the environment of sensor networks.

References

1. Caicedo, C.E., Joshi, J.B., Tuladhar, S.R.: IPv6 security challenges. Computer **2**, 36–42 (2009)
2. Choudhary, A.R., Sekelsky, A.: Securing IPv6 network infrastructure: A new security model. In: 2010 IEEE International Conference on Technologies for Homeland Security (HST), pp. 500–506) (2010)
3. Dahill, B., Levine, B.N., Royer, E., Shields, C.: A secure routing protocol for ad-hoc networks. Electrical Engineering and Computer Science, University of Michigan, Tech. Rep. UM-CS-2001–037 (2001)
4. Hu, L., Evans, D.: Using Directional Antennas to Prevent Wormhole Attacks. In: NDSS (2004)
5. Hu, Y.C., Perrig, A., Johnson, D.B.: Packet leashes: a defense against wormhole attacks in wireless networks. INFOCOM **2003**, 1976–1986 (2003)
6. Hu, Y.C., Perrig, A., Johnson, D.B.: Rushing attacks and defense in wireless ad hoc network routing protocols. In: Proceedings of the 2nd ACM Workshop on Wireless Security, pp. 30–40 (2003)
7. Khalil, I., Bagchi, S., Shroff, N.B.: MOBIWORP: Mitigation of the wormhole attack in mobile multihop wireless networks. Ad Hoc Netw. **6**(3), 344–362 (2008)
8. Papadimitratos, P., Haas, Z.J.: Secure routing for mobile ad hoc networks. In: The SCS Commnication Networks and Distributed Systems Modeling and Simulation Conference (CNDS), pp. 193–204 (2002)
9. Thread. http://threadgroup.org/About.aspx
10. Winter, T.: RPL: IPv6 routing protocol for low-power and lossy networks (2012). https://tools.ietf.org/html/rfc6550
11. Zigbee. http://www.zigbee.org/

Filtering Duplicated Location in Tracking Traffic Data

Swe Swe Aung and Thinn Thu Naing

Abstract Intelligent Transportation System (ITS) has been becoming an integral part of life in city, giving at the result of great impact by utilizing the communication, computing and sensor technologies to solve the relating problem of transportation such as traffic congestions. Traffic congestion is used to curse to citizen and an ongoing problem in almost urban areas. The purpose of this paper is mainly to provide the data without noises as much as possible to Traffic Detection System (TDS) based on GPS_enable Mobile phone. The system is constructed into two parts: Client side (Mobile device) and Cloud Backend Server. In this work, the process of transportation mode filtering is carried out on the Client side applying Moving Average Filtering method and then the filtering location duplicated data continues to work out on Server side based on the Client's result. In order to solve the server side issue, the distance based clustering method, OPTICS: Ordering Point To Identify the Clustering Structure, is mainly utilized. Afterward, the accuracy of the system is measured by Purity, F-measurement and Entropy method. To execute closeness between GPS points, the distance between them is measured by using Haversine Formula.

Keywords Cloud · GPS · ITS · TDS · Traffic · Transportation mode

1 Introduction

About the important of Intelligent Transportation System (ITS) has been introduce in abstract where and how. Moreover, it is also playing a vital role in the economy and its environment. ITS utilizes the communication, computing and sensor technologies to provide innovative services to the different transportation modes and to solve the

---

S.S. Aung(✉) · T.T. Naing
University of Computer Studies, Yangon, Myanmar
e-mail: {swesweaung.ucsy,thinnthu}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_33

relating problem of transportation such traffic congestion. Traffic congestion is an all familiar bane to citizen and a continuing problem in almost metropolitan areas. Traffic congestion increases travel time, air pollution, carbon dioxide (CO2) emissions and fuel use because cars cannot run efficiently. Therefore, traffic management system has got much attention from the research community.

The previous researchers had been mainly utilizing the network of fixed sensors, such as Cameras, road sensors on roadside and so on, for the traffic management system. One of the major draws is to cope with the cost of installation and continuous maintaining these sensors throughout the road network. Later, one popular and tremendous technique for intelligent transportation system is GPS_enable smart phone. The use of mobile device becoming to spread to the massive people very quickly both in the developed and developing countries can detect and learn the current situation of traffic condition more precisely and also effectively in deployment cost. However, traffic information could still be biased if the whole population is recognized as a useful data. Every motorize transport should not be recognized as a traffic prove. In a bus or taxi, almost passengers can have mobile phone.

Luckily, the sensors, such as GPS, camera, microphone, and accelerometer and so on, are an essential one for implementing new applications such as transportation, healthcare and so on. However, it still has limited resources such as battery life, network bandwidth, and storage capacity and processor performance. The limitations of Smartphone can be solved by using the advantages of Cloud Computing. The Fig. 1. shows the general architecture of the system.

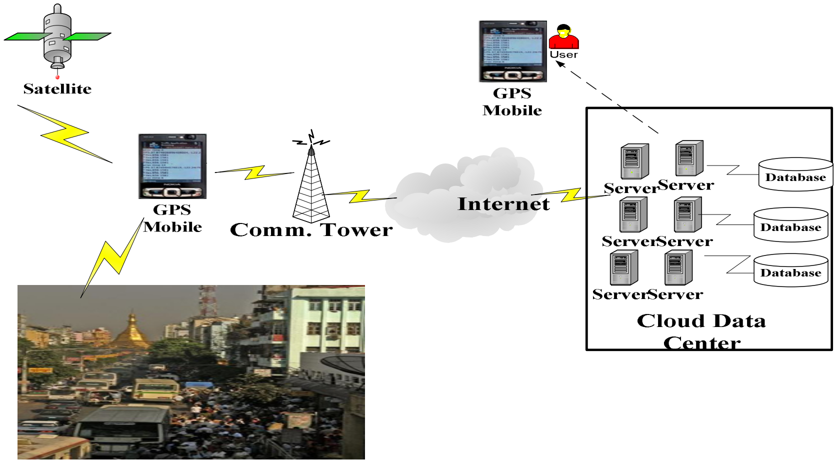


Fig. 1 Overall System Architecture

2 Design of System Architecture

On client side, this system focuses on analyzing whether the user is travelling with motorized transportation mode or non-motorized transportation mode by only using GPS and an accelerometer data.

For estimating the situation of traffic, it does not need to analyze the raw GPS data to get the detail information whether bike or car or motorbike. While bus, taxi, biking and motorbike are defined into the group of motorized transportation mode, only walking man is recognized as non-motorized mode in this system. This system does not take into account every GPS as traffic probe.

On the Server side process, it detects whether the submitted data comes from the same vehicle or not. If so, all data from the same bus is defined as only one traffic probe. Later, these probes are going to be used as a historical data to improve the quality of the system. The following figure gives the design of the system.

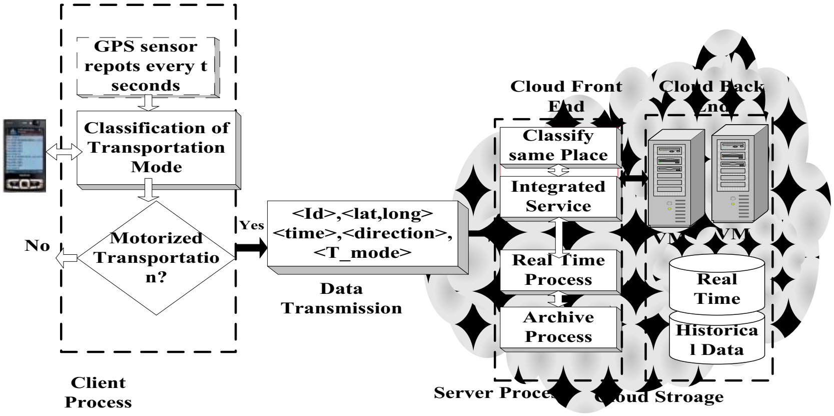

GPS sensor repots every t seconds
Classification of Transportation Mode
Motorized Transportatio n?
Yes
No
<Id>,<lat,long> <time>,<direction>, <T_mode>
Data Transmission
Cloud Front End
Cloud Back End
Classify same Place
Integrated Service
Real Time Process
Archive Process
VM
VM
Real Time
Historica l Data
Client Process
Server Process
Cloud Stroage

Fig. 2 Design of System Architecture

3 Related Work

D. Patterson, et al. [1] and L. Liao, et al. [2] used an unsupervised learning method to identify the transportation mode of a traveler. The transportation modes that were detected in this system include buses, cars and walk. The system was able to estimate a travelers' goal. In addition to GPS data, the system also used the history data of users. Historical information includes, past user trips and information about where the users parked their cars.

N. D. Lane, et al. [3] also analyzed the user's current transportation mode by applying a Boosted Naïve Bayes classifier to data collected from the sensors of the mobile device for the analysis of traffic patterns and CO2 emissions. For generating this Boosted Naïve Bayes classifier, it needed to develop an application to collect training data firstly. Importantly, the propose approach was different than that in [2]. The system did not give bother to get the training data previously. It kept the true data from the result of mobile device as a historical data and afterward these history data are integrated with the live result to improve the accuracy.

[4] has a different focus; it considered extracting the semantic location from outdoor positioning systems. Likewise, Hightower [5] learnt and recognized the places a mobile user visited by observing the Wi-Fi and GSM radio fingerprints. This work did not consider Wi-Fi or GSM information.

4 Client System Model

4.1 *Preprocessing*

To illustrate the flow of filtering raw GPS data, this system needs to identify the following notations for system parameters.

Table 1 GPS sensor log descriptions.

Notation	Description
p	GPS point which includes latitude, longitude ,speed and time.
Pi	Set of GPS point(lat, lon, v, t) , pi ϵ p=(i=1,2,…,n)
lat	Latitude
lon	Longitude
t	The time stamp of the sensor report in seconds
v	The current ground speed of the device
h	The direction of travel
acc	The accuracy level of the latitude and longitude coordinates
tm	Transportation mode (walking)
nm	Non transportation mode (non-walking mode)
GPS log	A sequence of GPS points
R	<id, lat, lon, v, t, s, tm>
D	Threshold for calculating distance is 0.6 km

4.2 *Client Process Algorithm*

Assume it computes the average speed every 15 seconds.

Table 2 Client Process Algorithm

```
Client Process Algorithm
Input : GPS data= <id,lat,lon,v,t>
Output : R= <id,lat,lon,v,t,s,tm>
Begin
    Step 1.    Receipt p={p1,p2, . . . , pn} within a time window
    Step 2.    Average Speed (P.V)= (∑i=1 to n piv) / n
    Step 3.    Filtering speed using Moving Average Filtering Method
                   If(P.V > Vt)
    Step 4.            Submit R to backend Server
                   Else
    Step 5.            Calculate distance (D) within five minutes
                   If(D>threshold)
    Step 6.            Submit R to backend Server
                   Else
    Step 7.            Go to step 1
End
```

The above algorithm is for computing the filtering process on the mobile device (Client side). Responsively it has to receipt the raw GPS data from the device and then filters out mode of transportation (driving or bus) by calculating these raw GPS data. In other word, it classifies that the mobile device carrier is taking bus or walking. However, there are some problems in analyzing the Transportation mode (walking or not). In real world, the buses short stopping at the any bus stop and at the traffic light is similar to walker mode. In order to solve this issue, Moving Average Filtering method is applied to smooth the raw and noisy GPS data. This method adjusts the targeted points by replacing the average speed of its neighbors. In equation form, this is written:

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i+j] \tag{1}$$

In this equation, $x[\ ]$ is the input signal, $y[\ ]$ is the output signal, and M is the number of points used in the moving average. Finally, the amount of noise still included in data is computed by taking Mean Square Error equation. In equation form,

$$\mathrm{MSE}(X) = E\left((X-\mu)^2\right) = (\sigma/\sqrt{n}) = (\sigma 2/n) \tag{2}$$

The following Fig. 3. and Fig. 4. compare the velocity before and after applying Moving Average Filtering method.

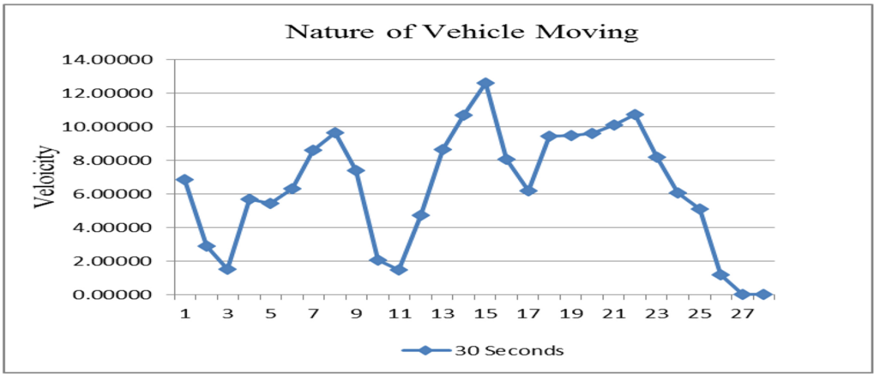


Fig. 3 Raw GPS data including Noise

Table 3 Estimation Errors in Analyzing Transportation Mode.

Total Number	**Statistics**				
	Mean	**Median**	**Standard Deviation**		**MSE**
47	5.278858	22.88817	3.941534		0.330547

The Table. 3. illustrates the estimation errors of analyzing Transportation mode over the data collected for nearly one hours along Pyay road in Yangon, Myanmar. The Fig. 4. shows that it is a clear noise one.

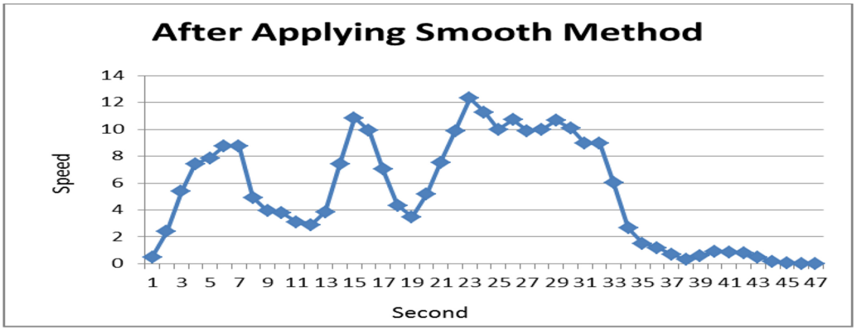


Fig. 4 Vehicle's velocity without Noise

5 Server Side System Model

5.1 Overall Architecture of Server Process

All data submitted form mobile device at every t time will be classified to define accurate traffic data. Firstly, these data are analyzed whether the corresponding devices are in the same place or not according the location, velocity and direction. In this case, this system needs to group the numerous GPS_location into similar counterpart. For this case. Ordering Point To Identify the Clustering Structure (OPTIC) is applied in this work. This algorithm operates on Dissimilarity matrix. The following figure shows the overall system architecture of Server side system.

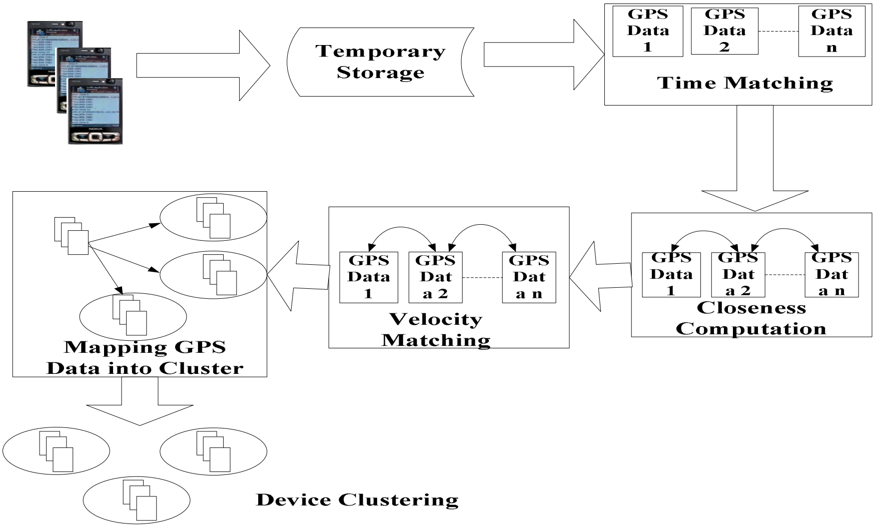

Temporary Storage
GPS Data 1
GPS Data 2
GPS Data n
Time Matching
GPS Data 1
GPS Dat a 2
GPS Dat a n
Closeness Computation
GPS Data 1
GPS Dat a 2
GPS Dat a n
Velocity Matching
Mapping GPS Data into Cluster
Device Clustering

Fig. 5 Server Side Clustering or Filtering Process Diagram

5.2 Server Side Algorithm

The Server side algorithm intends to cluster the GPS points into similar counterpart to reduce non usable data. The following table shows the process of Server Side Filtering Algorithm. In step1, the system receive Client GPS data and this received data is stored in the temporary table or Current Time Data Table. This table will be mainly accessed for the entire Clustering process. In step2, it firstly retrieves all the similar time GPS points from the table and then call the calculating distance function by passing these data as an argument. After this process, the function will return the distance matrix to the previous calling function. In step3, OPTICS Clustering function use this distance matrix to group GPS points that are closeness each other. The Table 5 shows the distance matrix based on real time data. The matrix computes the distance among devices. For example, $d(X_1, X_2)$ is a function that calculates the distance between X1 device and X_2 device and so on. In this calculation, function $d(X_n, X_n)$ accomplishes its task by using Haversine formula.

Table 4 Server Side Filtering Algorithm

```
Server Process Algorithm

Input : <id,lat,lon,v,t,s,tm>
Output : real time data result
Begin
    Step 1.    Receipt input <id,lat,lon,v,s,tm>
    Step 2.    Clustering Process
        1. Search already exist in the current time table
        2. If(exist) go to Step 3
             Else
                Calculate distance between the input GPS point
                and the previous data in the current table by using Haversine formula
        3. Use OPTICS algorithm to cluster points into similar location
    Step 3.    Store the result as a new one or update the old data
    Step 4.    Output real time data result
End
```

Where d(xi,xj) is the measured distance between points i and j. d(xi,xj) is a nonegative number that is close to 0 when two points i and j are near each other.

The distance between these devices is larger than 0.04 km because they are being on different vehicles. The two cluster can be seen clearly in the distance matrix Table 5. in red colored. The distance matrix Table 5 is the result of using the Cluster function to group the same counterpart according to their distance. In clustering process, the OPTICS clustering method is used in this work. Before clustering, the core object is defined by selecting the smallest distance between them.

Table 5 Distance Matrix Table

Column1	X1	X2	X3	X4	X5	X6
X1	0 km	0.0 km	0.0 km	0.2 km	0.2 km	0.2 km
X2	0.0 km	0 km	0.0 km	0.2 km	0.2 km	0.2 km
X3	0.0 km	0.0 km	0 km	0.2 km	0.2 km	0.2 km
X4	0.2 km	0.2 km	0.2 km	0 km	0.0 km	0.0 km
X5	0.2 km	0.2 km	0.2 km	0 km	0 km	0.0 km
X6	0.2 km	0.2 km	0.2 km	0.0 km	0.0 km	0 km

If all distances are identical, one of them is randomly chosen. The core object must has at least minimum points (MinPts) in its neighborhood. The MinPts is defined 3. After that it also computes the core distance and reachability distance using this distance matrix. The Fig. 6. indications the final outcome result of Clustering Algorithm.

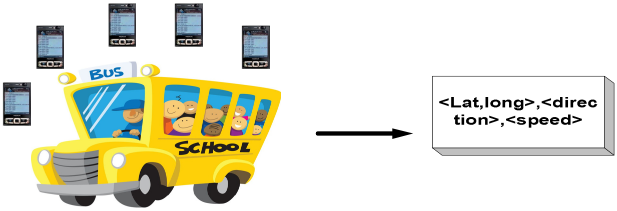


Fig. 6 Clustering Mobile Devices into Same Group

5.3 Evaluation of Server Side Process

The system finds out the cluster every 15 seconds by grouping same distance devices and also similar velocity. According to the result, the system well performs when the vehicle moving high speed even the velocity approximate to walking speed. However, if the vehicle is constantly stopping at junction for long time, the precision and F-measure will decrease and entropy value will increase. But the system can still tolerance and be allowable this state as it always checks back the previous state.

The performance analysis of the system is done by using three measurements, Purity, F-measurement and Entropy. Fig. 7. matches the precision of two clusters. While the value of two clusters between second 60 an 90 falls little, at the other time the two clusters are identical. It also means that they can be clustered well. The Fig. 8. compares the entropy valued of two clusters. If the entropy value approaches to 0, the cluster is in good condition. At almost time, the two clusters are identical because of their good purity. In other word, it is concluded that they are on the individual bus.

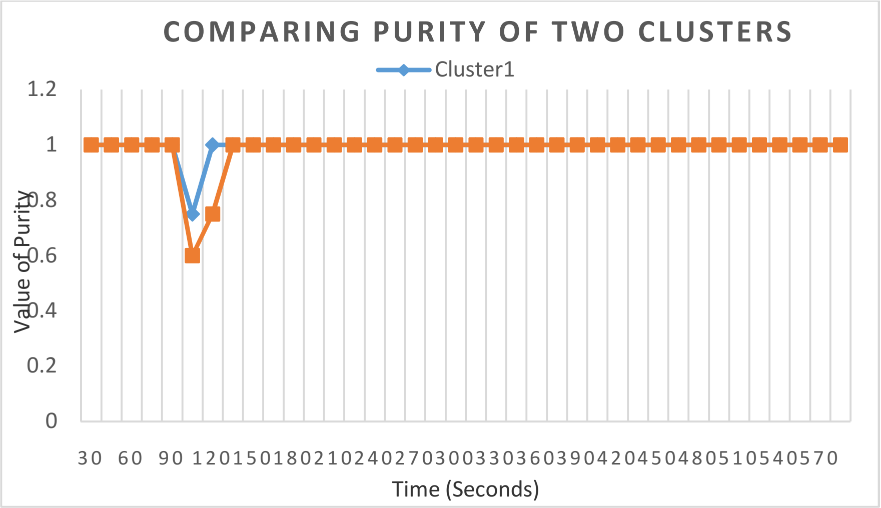


Fig. 7 Purity Measurement of Two Clusters

The Fig. 9. illustrations the F-Measurement of two clusters. For a perfect clustering, the maximum value of the F-measure is one. The value of tow clusters is a little different between 90 and 120 second. Nevertheless the system wholly overcome the impact of the short distance stopping because it recognizes the previous state when the vehicle has to stop at the bus stop for short time.

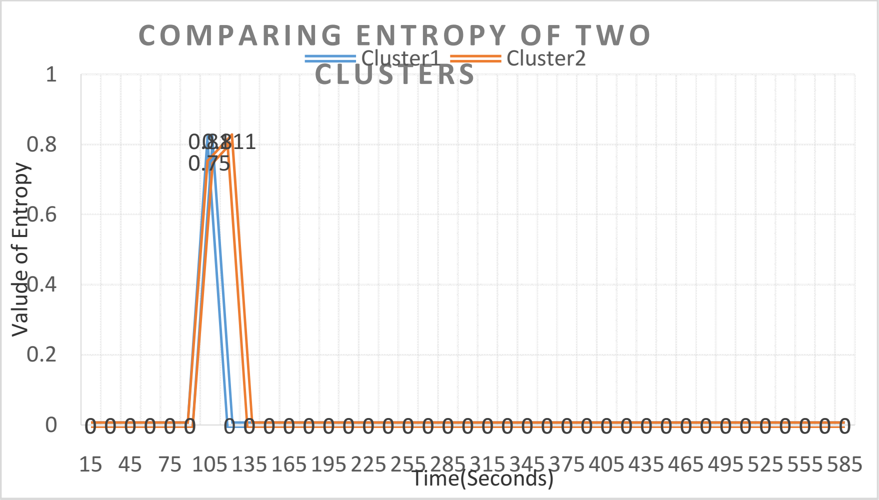


Fig. 8 Entropy Measurement of Two Clusters

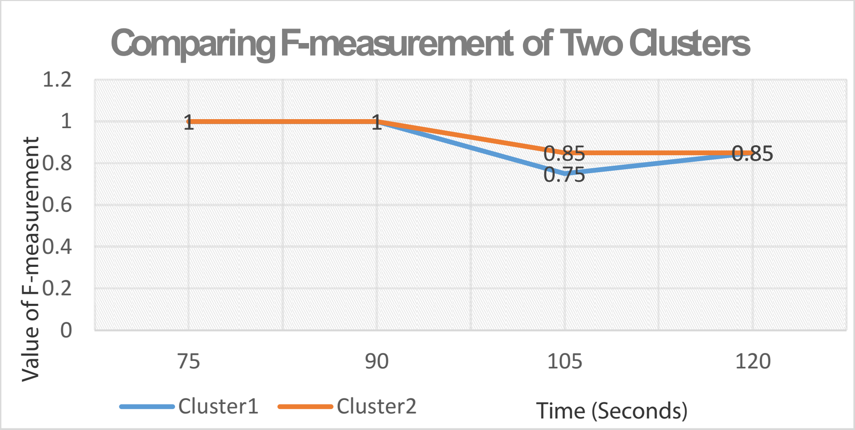


Fig. 9 F-Measurement of Two Clusters

6 Conclusion

In this paper, we highlight tracking Traffic data for (TDS) using GPS_enable Mobile phone aiming at providing more accurate data. For this reason, the system firstly filter the noisy GPS data into Transportation mode or non-Transportation mode. Afterwards all incoming GPS logs are not taken into account as the true Traffic data for Traffic system. Some of them may be on the same bus. Therefore, the duplicated locations are removed out by applying clustering method. From the standpoint of Traffic Management System, this approach may provide more accurate and efficient data to Traffic System. Moreover, Using Mobile GPS approach becomes at most no installation and maintenance cost compared with others.

References

1. Patterson, D., Liao, L., Fox, D., Kautz, H.: Inferring High-Level Behavior from Low-Level Sensors. In: ACM UBICOMP 2003 (2003)
2. Liao, L., Fox, D., Kautz, H.: Learning and Inferring Transportation Routines. In: AAAI 2004 (2004)
3. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of Mobile Phone Sensing (September 2010)
4. Liu, J., Wolfson, O., Yin, H.: Extracting Semantic Location from Outdoor Positioning Systems. In: Int. Workshop on Managing Context Information and Semantics in Mobile Environments (MCISME) (2006)
5. Hightower, J., Consolvo, S., LaMarca, A., Smith, I., Hughes, J.: Learning and Recognizing the Places We Go. In: ACM Conference on Ubiquitous Computing (2005)

Finding Shortest Path and Transit Nodes in Public Transportation System

Myat Thu Zar and Myint Myint Sein

Abstract One of the most important pieces of information to be delivered to public transportation users is the quickest bus routes between user specified current and destination. In this system, A* algorithm is used to find the shortest path and bus route information in Yangon downtown area. Not only direct information but also transit information can give between current and destination bus stop. Bus routes, bus information, bus stops and their related position, latitude/longitude and road names are stored in spatial database. When the users want to know the bus route information, firstly they give their current location and desired destination to the system and then the related bus routes are retrieved from database. After getting the bus routes, A* algorithm is used to calculate the shortest path. By using the TRANSIT algorithm, transit route and information will be displayed on the map. Vector map of Yangon region is used to create as base map in this paper.

Keywords A* algorithm · TRANSIT algorithm · Road network · Public transportation system

1 Introduction

The traffic problem of Yangon area has been directly affects the quality of the citizens of the area. Most of citizens in Yangon area make use of public transportation such as bus, taxi, rail and so on. Figure 1 shows some types of public transport in Yangon region. The effects of the traffic of Yangon area are the pollution in every aspect such as air, sound, the time wasting on the road and also the hydrocarbon fuel. Even though the numbers of the roads are increasing, the traffic problem in Yangon area is still occurring. Advanced public transportation systems are an important component of smart transportation systems for reducing traffic demands. We report algorithms for providing bus information to tourists and

---

M.T. Zar(✉) · M.M. Sein(✉)
University of Computer Studies, Yangon, Myanmar
{myatthuzar86,myintucsy}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_34

commuter to encourage people to use public transportation systems. Providing transfer information is another effective measure for promoting the public transportation systems [4].

There are a number of algorithms that can be used to determine shortest route between two nodes in a network. Among them Dijkstra's algorithm and A*algorithm are more efficient. Dijkstra's algorithm determines the shortest route between the source node and every other node and A* algorithm much like Dijkstra based on heuristics strategy [8].

An efficient public transport system, that integrates residential and employment hubs, plays an important role in moving skills, labor and knowledge within and between markets, stimulating productivity [6]. Public transport corridors have shown to create activity centers that encourage and foster transit orientated development, and in turn through providing easy access to public transport increase adjacent property values.

In some online mapping service, the user may have seen satellite imagery. When these images are captured from a satellite or an airplane, they are just plain images, like photographs. But to display these images on a map, they need to be associated with map coordinates. This process is called GeoReferencing. Once the image is associated with the map coordinates it can be overlaid on top of street maps. For georeferencing, some user can use GIS software such as ArcGIS or QGIS to georeferencing and otherwise un-referenced image or scanned map, and load them into Oracle spatial [3].

The remainder of the present paper is organized as follows. Section 2 describes the basic representative network structure. In Section 3 describes the applied method A* algorithm and related function for this system. Section 3 briefly explains about system overview and data. Section V discuss about how to do experiment the proposed system on Yangon downtown region. Finally section 6 summarizes present paper and describes future plan.

Fig. 1 Some types of public transport in Yangon region

2 Road Network Representation

A public road network is composed by some nodes (bus stop) in this paper, the links connecting two nodes and bus lines. Define a public road network as G, G={N,E,R}, where N={1 ≤ i ≤ n} denotes the set of all nodes, and n is the number of nodes; the origin node and the destination node is O, D respectively. E={1 ≤ e ≤ m} is the set of all transit links, and m is the number of links; R={ 1 ≤ r ≤ u } is the set of all bus lines and u is the number of links. In this paper, algorithm employs two functions that relate bus stops and bus routes. We use the distance on the road network as the cost in this paper. Figure (2) shows the example road network of Yangon region.

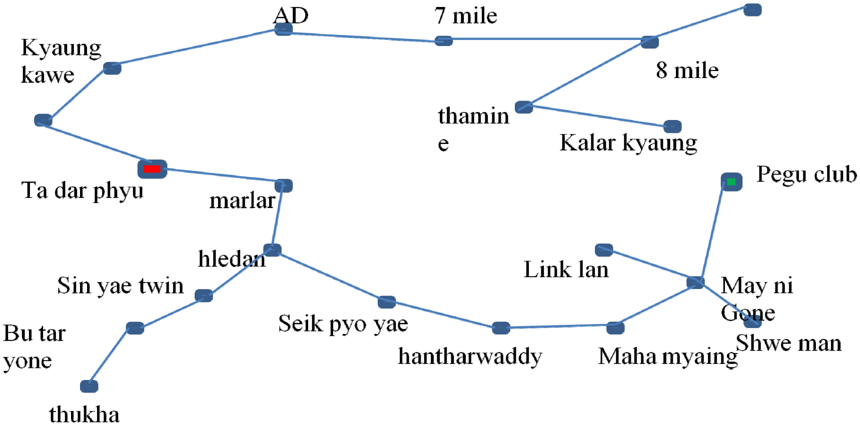
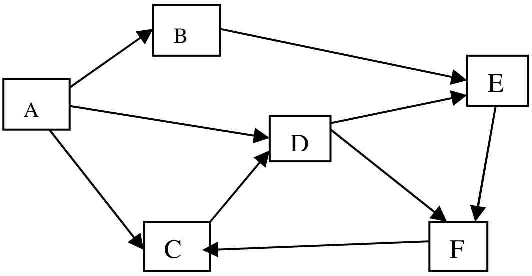


Fig. 2 Sample road network of Yangon region

2.1 Adjacency Matrices and Distance Matrices

In this paper, we represent the connectivity among locations in transportation networks with adjacency matrices [1]. We designate each location in the network a unique number. The value of a cell $M_{i,j}$ of an adjacency matrix M is set to the number of direct ways that we may travel from a location i to j. The following example, figure 3 illustrates how we represent a simple road network with an adjacency matrix T and distance matrix

Given a weighted graph that represents a street network, a problem instance consists of a number of sources and destinations located in the graph [9]. For these sets of nodes we want to know the distances from all sources to all destinations. Hence the result of such a M×N query is a matrix of distances. An entry of this distance table denotes the distance from the source corresponding to the current column, to the target corresponding to the current row.



$$T= \begin{array}{c|cccccc|} & A & B & C & D & E & \\ A & 0 & 1 & 1 & 1 & 0 & 0 \\ B & 0 & 0 & 0 & 0 & 1 & 0 \\ C & 0 & 0 & 0 & 1 & 0 & 0 \\ D & 0 & 0 & 0 & 0 & 1 & 1 \\ E & 0 & 0 & 0 & 0 & 0 & 1 \\ F & 0 & 0 & 1 & 0 & 0 & 0 \end{array} \qquad \begin{array}{c|cccccc|} & A & B & C & D & E & \\ A & 0 & 1 & 1 & 1 & \infty & \infty \\ B & \infty & 0 & \infty & \infty & 1 & \infty \\ C & \infty & \infty & 0 & 1 & \infty & \infty \\ D & \infty & \infty & \infty & 0 & 1 & 1 \\ E & \infty & \infty & \infty & \infty & 0 & 1 \\ F & \infty & \infty & 1 & \infty & \infty & 0 \end{array}$$

Fig. 3 Example bus route and its adjacency matrix and distance matrix

3 System Methodology

3.1 A Algorithm*

A* algorithm is a graph search algorithm that find a path from a given initial node to a given goal node in a mapped area. The A* algorithm by Hart and Nilsson [5] formalized the concept of integrating a heuristic into a search procedure. It employs a 'heuristic estimate' h (n) that gives an estimate of the best route that goes through the node. It visits the nodes in order of this heuristic estimate. It follows the approach of best first search [7].

A* algorithm are the building of a "closed list" to record areas already evaluated, an "OPEN list" to record areas adjacent to those already evaluated, and the calculation of distance traveled from the "start point" with estimated distance to the "goal point". The heuristic used to evaluate distances in A* is:

$$f(n)=g(n) + h(n) \tag{1}$$

Where -g(n) is the cost of the path from the starting node to any node n, and
-h(n) is the heuristic estimated cost from any node n to the goal

3.2 A Algorithm Steps*

1. Create a search graph, G, consisting solely of the start node, n_0. Put n_0 on a list called OPEN.
2. 2. Create a list called CLOSED that is initially empty.

3. If OPEN is empty, exit with failure.
4. Select the first node on OPEN, remove it from OPEN, and put it on CLOSED. Called this node is n.
5. If n is a goal node, exit successfully with the solution obtained by tracing a path along the pointers from n_0 to n_n in G. (The pointers define a search tree are established in step 7).
6. Expand node n, generating the set, M, of its successors that are not already ancestors of n in G. Install these members of M as successors of n in G.
7. Establish a pointer to n from each of those members of M that were not already in G (i.e., not already on either OPEN or CLOSED). Add these members of M to OPEN. For each member, m, of M that was already on OPEN or CLOSED, redirect its pointer to n if the best path to m found so far is through n. For each member of M already on CLOSED, redirect the pointers of each of its descendants in G so that they point backward along the best path found so far to these descendants.
8. Record the list OPEN in order of increasing f(n) values
9. Go to step 3.

3.3 Heuristic Function

Heuristic search has been widely used in both deterministic and probabilistic planning. The heuristic function can be used to control A*'s behavior. Euclidean distance is a common method for h(n).

$$\text{Distance}=\text{sqrt}(dx^2+dy^2) \tag{2}$$

- x_2 = coordinate of the goal location
- x_1=coordinate of the current location
- y_2=coordinate of the goal location
- y_1=coordinate of the current location
- $dx=|x_2-x_1|$
- $dy=|y_2-y_1|$

If h(n) is always lower than to the cost of moving from n to the goal, then A* is guaranteed to find a shortest path. If h(n) is exactly equal to the cost of most of moving from n to the goal, then A* only follow the best path and never expand anything else, making it very fast. If h(n) is sometimes greater than the cost of moving from n to the goal, then A* is not guaranteed to find a shortest path, but it can run faster. At the other extreme, if h(n) is very high relative to g(n), then only h(n) plays a role and A* turns into Best-First-Search.

3.4 Transit Algorithm

The transit algorithm is one of the methods for finding shortest paths in very large road networks. One of the advantages of TRANSIT is that the final query time is completely indifferent to the path length, but only depends on the number of transit nodes of the origin and destination [13], [14]. Figure 4 show an example for considering transit node in bus route. Assume there are three bus services: A→B→C→D→F, H→B→C→D→E and G→D→E. The people want to travel from A to E. Clearly, at some point have to change a service and it can be either at B, C and D. On the other hand, potentially may change a line at B or D. Let $G_S=(V_S,E_S)$ to be a bus station. Denote Si be a set of all services that depart from bus stops $s_i \in V_s$

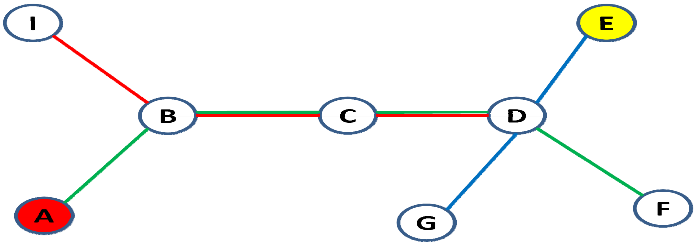


Fig. 4 Example of transit node in bus route

3.4.1 Algorithm for Identifying Transit Nodes

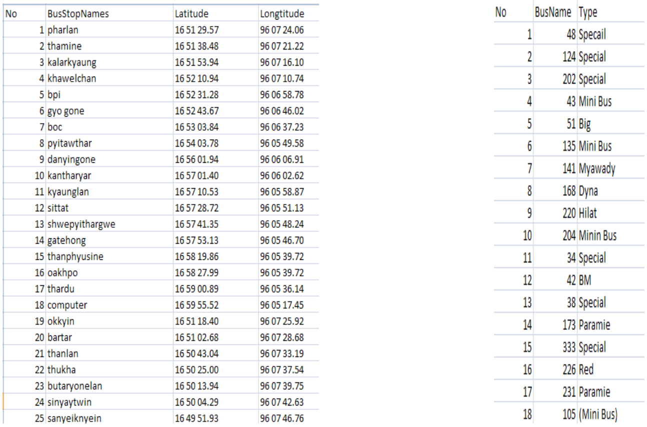
```
for $s_i$ Є $V_S$
        for (($s_i$,$s_j$ ) Є E)
        if  ($S_j$ C $S_i$)
                H       :=H \ $s_j$
                H       := $s_i$
        else if ($S_i$ C $S_j$)
                H       :=H \ si
                H       := $s_j$
        else if ($S_i \neq S_j$)
                H       :=$s_i$
                H       :=$s_j$
```

Where,

S_i ,S_j = a set of bus lines
H= transit node
V=vertex(bus stop)
E =edge (bus link)

4 System Overview

In this paper, the bus public transportation information for Yangon downtown region such as bus stops name, bus routes name, bus type, latitude and longitude position are stored in spatial database. Time complexity in A* algorithm is O(n log n), n is the number of node. This system can be used without internet



No	BusStopNames	Latitude	Longtitude
1	pharlan	16 51 29.57	96 07 24.06
2	thamine	16 51 38.48	96 07 21.22
3	kalarkyaung	16 51 53.94	96 07 16.10
4	khawelchan	16 52 10.94	96 07 10.74
5	bpi	16 52 31.28	96 06 58.78
6	gyo gone	16 52 43.67	96 06 46.02
7	boc	16 53 03.84	96 06 37.23
8	pyitawthar	16 54 03.78	96 05 49.58
9	danyingone	16 56 01.94	96 06 06.91
10	kantharyar	16 57 01.40	96 06 02.62
11	kyaunglan	16 57 10.53	96 05 58.87
12	sittat	16 57 28.72	96 05 51.13
13	shwepyithargwe	16 57 41.35	96 05 48.24
14	gatehong	16 57 53.13	96 05 46.70
15	thanphyusine	16 58 19.86	96 05 39.72
16	oakhpo	16 58 27.99	96 05 39.72
17	thardu	16 59 00.89	96 05 36.14
18	computer	16 59 55.52	96 05 17.45
19	okkyin	16 51 18.40	96 07 25.92
20	bartar	16 51 02.68	96 07 28.68
21	thanlan	16 50 43.04	96 07 33.19
22	thukha	16 50 25.00	96 07 37.54
23	butaryonelan	16 50 13.94	96 07 39.75
24	sinyaytwin	16 50 04.29	96 07 42.63
25	sanyeiknyein	16 49 51.93	96 07 46.76

No	BusName	Type
1	48	Specail
2	124	Special
3	202	Special
4	43	Mini Bus
5	51	Big
6	135	Mini Bus
7	141	Myawady
8	168	Dyna
9	220	Hilat
10	204	Minin Bus
11	34	Special
12	42	BM
13	38	Special
14	173	Paramie
15	333	Special
16	226	Red
17	231	Paramie
18	105	(Mini Bus)

Fig. 5 Some data of bus stops and bus name

connection. In Figure 5, show the some data of bus stop and bus route. When the users enter the place name (bus stop name) in the search box and in return the map shows a marker at the place. When the user give the source and destination bus stops, the system search the available bus lines from database. And then display bus line on the map. In figure 6, the process of system shows flow diagram. The process of associating a place name with coordinates on the map is called geocoding. In a spatial database, this data done as a line layer with the name of the place as an attribute to the location. This is one way of geocoding [11]. For addresses, the associated coordinates are not saved in a database directly, but computed using a method called linear referencing. The start and end addresses along a long segment are saved and intermediate address are interpolated and the coordinates are calculated.

5 Experiment

In this paper, the proposed system is tested on Yangon down town region, road network. Collect the bus routes information, bus stops, road name and their related geographical information data such as latitude/longitude and position in Yangon downtown area are collected from Google earth and use GPS GARMIN etrex-10 device. Yangon down town region is limited from left latitude 17 · 03' 56.63" N to right longitude 96· 19' 04.15"E.These collecting data are stored in spatial database which will use to develop the A-star algorithm of graph theory to calculate the shortest route. The system will be able to show the data to the user to select the optimal route and bus number with information of public bus transport for downtown region in Yangon. TRANSIT algorithm can be calculated for transit node in bus routes. Figure 7 show the Yangon region vector map for this system.

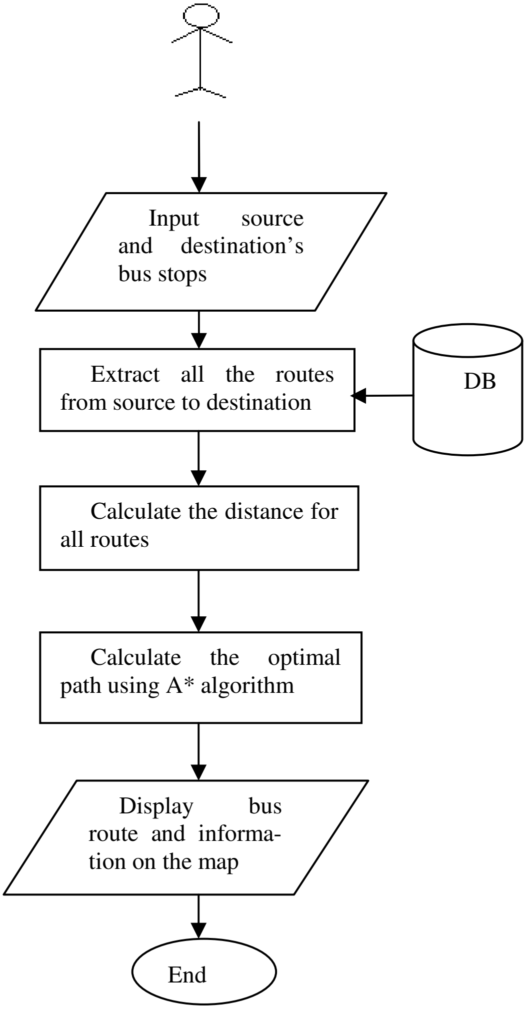

Input source and destination's bus stops
Extract all the routes from source to destination
DB
Calculate the distance for all routes
Calculate the optimal path using A* algorithm
Display bus route and information on the map
End

Fig. 6 System Flow Diagrams

The basic algorithm in shortest path computation is Dijkstra's algorithm. In practice, A* algorithm is used as a more efficient alternative because it uses heuristic function to guide the search toward the destination. A* is faster as compare to Dijkstra's algorithm because it uses Best First Search whereas Dijkstra's uses Greedy Best First Search. The major disadvantage of Dijkstra's algorithm is the fact that it does a blind search there by consuming a lot of time waste of necessary resources.

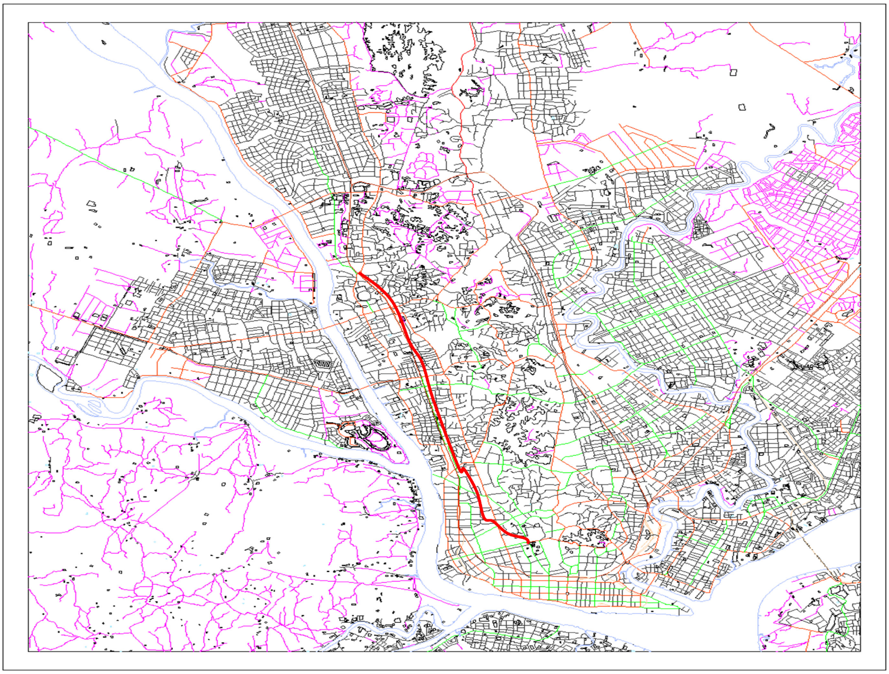


Fig. 7 Vector map of Yangon area

6 Conclusion

In this system, we presented an approach for finding shortest path based on distance in public transportation network. We presented an experimental results using Yangon area of public transportation. Public bus transportation system by A* algorithm apply in Yangon downtown area will assist the bus transportation user to the information and data which is essential for them and also provide the planning system to plan the route and transportation mode with the fare. A* is better when we know both starting point and destination point. A* is both complete (finds a path if one exists) and optimal (always finds the shortest path) if the user use an Admissible heuristic function. As a future work, the time or other factors such as waiting time at the bus stop, walking time to the bus stop will be used as weight for road network in Yangon area.

References

1. Pun-Cheng, L.S.C.: An interactive web based public transport enquiry system with real time optimal route computation. IEEE Transactions on Intelligent Transportation Systems **13**(2) (June 2012)
2. Transactions on Intelligent Transportation Systems **13**(2) (June 2012)

3. Pucher, J., Korattyswaroopam, N.: Rutgers University Neenu Ittyerah. The Crisis of Public Transport in India: Overwhelming Needs but Limited Resources **7**(3) (2004)
4. Samet, H.: University of Maryland,USA. Database and Geographical Information System (GIS) (2002)
5. Liu, C.-L., Pai, T.-W., Chang, C.-T., Hsieh, C.-M.: Path-Planning Algorithms for Public Transportation Systems. In: The 4th International IEEE Conference on Intelligent Transportation System, USA (August 2001)
6. Dai, L.: Fast Shortest Path Algorithm for Road Network and Implementation. Carleton University if Computer
7. Science COMP 4905 HONOURS PROJECT Fall Term (2005); Subadra, N., Bhagavan Das, M., Rama Seshagiri Rao, C.: Directed Graph Algorithms for Tours–A Case Study. Scholarlink Research Institute Journals (2011) (ISSN: 2141–7016)
8. Abhishek, G., Prateek, M., Rishabh, L., Ms, N.S.: Path Finding: A* or DIJKSTRA'S? IJITE 2(1) (January 2013)
9. Goldberg, A.V., Harrelson, C.: Computing the Shortest Path:A_ Search Meets Graph Theory, Technical Report MSR-TR-2004–24
10. Sebastian, K., Peter, S., Dominik, S., Frank, S.: Fast Computation of Distance Tables using Highway Hierarchies (July 25, 2006)
11. http://wiki.openstreetmap.org/wiki/Offline_Openstreetmap
12. http://gis.stackexchange.com/questions/28593/what-is-the-difference-between-geocoding-and-georeferencing
13. Bast, H., Funke, S., Matijievic, D.: Transit ultrafast shortest-path queries with linear-time preprocessing. In: 9th DIMACS Implementation Challenge (2006)
14. Bast, H., Funke, S., Matijevic, D., Sanders, P., Schultes, D.: In transit to constant time shortest path queires in road networks. In: Proc. 9th Workshop on Algorithm Engineering and Experimentation (ALENEX) (2007)

Modified Priority Search Tree for Location Aware Extracting Services

Linn Linn Phway and Myint Myint Sein

Abstract In this system, the data structure of Priority Search Tree is modified to query the data-points according to the user's defining range. A query in this new data structure returns all points in circular and rectangular range around a point with user preference distance to extract accurate services according to their geographic coordinates for GIS. A circular range query has the form query (q, r), and a rectangular range searching based on the envelope. The proposed range searching is the querying of all data-points P which are the subset of dataset R. After the specific information (e.g.; ATM, Hospital, Hotel, Shopping Center… etc.,) is extracted by the preference range, their coordinate are converted to image coordinate and show on the own map. When creating a map, sub images are extract from own database by their geographic coordinate. The proposed both for services database and sub-images database is based on the priority search tree (PST) developed by Edvard T. McCreight in the mid eighties.

Keywords Priority search tree · Range searching · LBS

1 Introduction

Location-based service (LBS) is the generation of Geographic Information System (GIS) and proposed system attempts to provide LBS. A location-based service can be defined as an information service provided by a device that knows where it is, and what the services are around it. Location-based services aim to provide new services to their users based on the knowledge of their locations. A range search is a query where the query key is a range. Searching for a range leads to more than one value is being reported. Range Searching system allows user to search for businesses and other requriement services that satisfy the user define geographical constraint. The areas of computer science where the data is geometric computer graphics such as GIS and spatial data structures are important. The data structures store a set ***R*** of geometric objects such that various queries can be answered efficiently. The most widely used queries are range queries, which request for all objects in ***R*** in a query

---

L.L. Phway(✉) · M.M. Sein
University of Computer Studies, Yangon, Myanmar
e-mail: {linlinphway,myintucsy}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_35

range ***Q***. Another important usage queries is nearest-neighbor queries, where one asks for the object from ***R*** closest to a query point [4].

Extracting Services System allow a user to search for businesses and other requriement services that satisfy a given geographical constraint. This system is developed in user defining range that will figure out the user location and searching the desired services around user.

Combining the functionally of map, location-detection devices, personal handheld devices, and database systems has resulted in the realizition of location-based services as commercial. There is necessary to defined specific location range so the user can decide to choose their destination more easily and quickly. Displaying the map is very easy to understand for users.

In this system, Global Positioning System (GPS) device is used to define and collect geographic coordinates for all places on earth. Download these points onto the computer and integrate them with mapping programs. The own Data-Set is created for Yangon region which contain the services, popular places, Latitude, Longitude, Position and Sub-Images. Each sub-image (approximately 0.15 km2) is grabbed from Google Earth and they have own min/max latitude and longitude according their size. Both for services database and sub-images database is developed based on the modified PST and range searching algorithm are considered. In this system, combination process of the sub images to achieve the map which involve the user's current location and desire service location according to the user define range. Coordinate converting Algorithm is presented for converting the geographic coordinate to image coordinate system and then the services are displayed on the map.

2 Related Works

Haibo Hu and Dik Lun Lee (2006) consider the (hyper)-rectangles range and propose the method for efficient in-memory processing and secondary memory pruning techniques for Range Nearest Neighbor (RNN) queries in 2D and high-dimensional spaces. Their techniques are generalized for k Range Nearest Neighbor (kRNN) queries for getting the k-nearest neighbors for every point in the defining range. Their idea is to speed up any type of Nearest Neighbor (NN) query by supporting with auxiliary solution-based index EXO-tree [2]. Sequential Scan, Projection, Cells, Kd Trees, Range Trees, k-ranges and Quad Tree for range searching are listed in [3]. They provide not only the methods for attacking multi-key searching problems and also display the data structures available for solving the particular problem of range searching [3]. The min-max priority search tree is the combination of a binary search tree and a min-max heap. In [5] show all the standard queries which can be done with a single min-max priority search tree instead of two separate versions of a priority search tree. The min-max priority search tree algorithm to enumerate all maximal empty axis-parallel rectangles amongst points in a rectangular region R in R^2 in $O(m \log n)$ time with $O(1)$ extra-space, where m is the total number of maximal empty rectangles.

Robin Y. Flatland, Charles V. Stewart (2000) a new data structure is introduced that obtain efficient overall query times by performing novel searches of multiple range trees and extending *k-nn* trees. Their search techniques and data structures generalize to algorithms for extending fixed polytope range queries and extending k-nn using polytope distance functions.

When queries *Θ(N)* points or require $E = \Omega(N)$ extensions with other algorithms in two dimensions, the overall retrieval time of their algorithms is $O(E+N)$. Their research give efficient algorithms for two instances of extending neighborhood problems stated as extending orthogonal range reporting and extending L_∞ k-nearest neighbors [6].Sunil Arya, David M. Mount (2006) show that given a bounded range Q of diameter ω and $\epsilon < 0$, an approximate range query meaning that points lying within distance $\omega\epsilon$ of the boundary of Q either may or may not be counted. They show that in any fixed dimension d, a set of n points in R^d can be preprocessed in $O(n\log n)$ time and $O(n)$ space, such that approximate queries can be answered in $O(\log n+(1/\epsilon)^d)$ time. They also present a lower bound for approximate range searching based on partition trees of $\Omega(\log n+(1/\epsilon)^{d-1})$, which implies optimality for convex ranges [7].

3 Priority Search Tree

The Priority Search Tree is a data structure used for performing semi-infinite range queries. Priority search tree algorithms can also be used to enumerating all intersecting pairs of rectangles. Priority search tree, for representing a dynamic set D of ordered pairs [x,y] over the set 0,1,…,k-1 of integers. Priority search tree have basically two types;(i) Radix Priority Search Tree and (ii)Balanced Priority Search Tree. In our propose system radix priority search tree is emphasized [1]. The first invariant is a priority queue invariant on y-values. It asserts that for any node t in a radix priority search tree, if t.left is not NIL then t.p.y ≤ t.left ↑.p.y, and if t.right is not NIL then t.p.y ≤ t.right ↑ p.y. This first invariant constrains only direct ancestor-descendant relations; it does not constrain sibling relations at all [1].The second invariant is a radix search tree invariant on x-values. It asserts that associated with each node t in a radix priority search tree is an x-interval [lower, upper] within which t.p.x lies. The x-interval associated with the root of the radix search tree is the interval KeyBound. For any node t such that t.left is not NIL, the x-interval associated with the node t.left ↑ is [lower .. floor(lower+upper)/2)). For any node t such that t.right is not NIL, the x-interval associated with the node t.right ↑ is [floor((lower + upper)/2)..upper). The y-value of new pair smaller than the root, the new pair "beats" the already root pair in InsertPair procedure. If not, then the new pair is inserted recursively into either the left or right sub-tree, determined by its x-value[1].

4 Constructing Modified PST for Services Spatial Database

In my proposed system, the first invariant is the same as radix priority search tree. However, in the second invariant is the tree invariant on distance between data points. It is useful for both rectangular and circular range search. The y-value of new pair smaller than the root, the new pair "beats" the already root pair in modified PST. If not, then the new pair is inserted recursively into either the left or right sub-tree, determined by its distance from root node.

Table 1 Sample of Storage Dataset

Place Name	*Service Type*	*Image Number*	*Latitude(N)*	*Longitude(E)*
Shwe Dagon Pagoda	Pagoda	10132	16 47 54.37	96 8 58.07
Parami Children Hospital	Hospital	8158	16 50 59.93	96 9 22.60
Dagon Centre-1	Shopping Centre	9920	16 48 14.49	96 8 15.37
Hel Pin Bus stop	Bus stop	10337	16 47 28.07	96 8 27.23
Kanbawza Bank	Bank	10452	16 47 17.75	96 10 14.20

Modified Priority Search Tree

```
Modified PST : ( t, p, Min dist, Max dist)
Input : $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ in R;
Output ; The Priority Search Tree of those points in R
VAR
        mPST : t;
        p : Pair;
        Min dist : 0;
        Max dist : $\sqrt{(\max X - \min X)^2 + (\max Y - \min Y)^2}$;
        newPr : $(x_n, y_n)$;
        $p_d$ = dist(p, t.p)
BEGIN
    IF t = NIL THEN
        BEGIN
                NEW(t); (*add a pair from R to root node*)
                t.p := newPr;
                t.left := NIL;
                t.right := NIL;
        END
    ELSE IF t.p.x <> newPr.x(*assumes unique x values*), THEN
        BEGIN
                IF newPr.y < t.p.y THEN(*new pair beats esisting one*)
                        BEGIN
                                p := t.p; t.p := newPr
                        END(*exchange new/existing*)
                ELSE
                        p := newPr;
                        Midium := (Min dist + Max dist) DIV 2;
                                IF $p_d$ < Midium Dist, THEN
                                        modified PST( t.left, p, Min dist, Midium dist)
                                ELSE
                                        modified PST( t.right, p, Midium dist, Max dist);
                                END;
                END;
        END;
   END;
END;
```

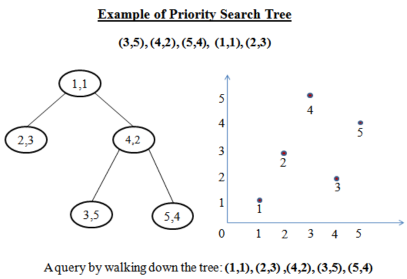

Example of Priority Search Tree
(3,5), (4,2), (5,4), (1,1), (2,3)
1,1
2,3
4,2
3,5
5,4
0 1 2 3 4 5
1 2 3 4 5
A query by walking down the tree: (1,1), (2,3) ,(4,2), (3,5), (5,4)

Fig. 1 Example of PST

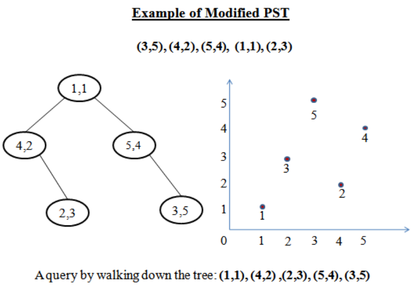

Example of Modified PST
(3,5), (4,2), (5,4), (1,1), (2,3)
1,1
4,2
5,4
2,3
3,5
0 1 2 3 4 5
1 2 3 4 5
A query by walking down the tree: (1,1), (4,2) ,(2,3), (5,4), (3,5)

Fig. 2 Example of Modified PST

5 Range Searching in MPST

In this system, the range searching is developed by using modified priority search tree. Vincenty's formulae are two related iterative methods used in geodesy to calculate the distance between two points on the surface of a spheroid, developed by Thaddeus Vincenty (1975). The first (direct) method computes the location of a point which is a given distance and azimuth (direction) from another point. The second (inverse) method computes the geographical distance and azimuth between two given points. They have been widely used in geodesy because they are accurate to within 0.5 mm (0.020″) on the Earth ellipsoid [9].

5.1 Rectangular Range Search

For Rectangular Range Search, we use Vincenty's direct method to determine the maximum and minimum bounded that cover for the user requirement range. Latitude and Longitude of user current location as lat_1, lon_1, user's define range is used as s and α have four variable such as 0°, 90°, 180° and 270°. In geographic coordinate system, true north is 0° and it can get the maximum latitude, in 90° can get maximum longitude, in 180° can get minimum latitude and 270° can get minimum longitude. After processing this step, we get the two pairs of latitude and longitude which two pair are cover the user define range that is Max/Min Bounding Rectangle [8].

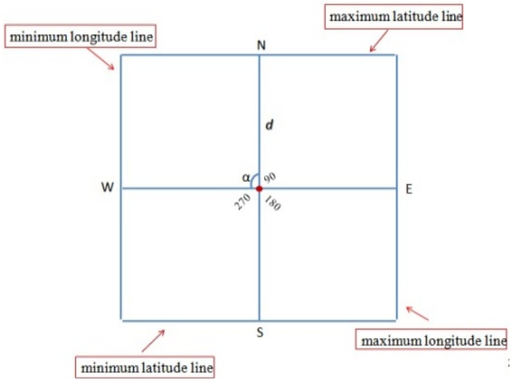

maximum latitude line
minimum longitude line
N
d
α
90
W
E
270
180
S
maximum longitude line
minimum latitude line

Fig. 3 Defining Max/Min Bound

```
Rectangular Range Search in modified PST
RectangularRangeSearch(t : MPST, $q, r$)
VAR
        continue : BOOLEAN;
        q = query point;
        r = range(distance);
        ($x_{min}$, $x_{max}$, $y_{min}$, $y_{max}$) := Vincenty's Direct Method(q, r,0°,90°,180°,270°);
BEGIN
        IF t ≠ NIL, THEN
                IF ($y_{min} \le$ t.p.y $\le y_{max}$) AND ($x_{min} \le$ t.p.x $\le x_{max}$)THEN
                        continue := Report(t.p);
                ELSE IF t.p ≠ leaf, THEN
                        RectangularRangeSearch(t.p.left, $x_{min}$, $x_{max}$, $y_{min}$, $y_{max}$, Report);
                        continue := Report(t.p.left);
_               ELSE IF t.p.left = leaf t.p.right ≠ NIL, THEN
                        RectangularRangeSearch(t.p.right, $x_{min}$, $x_{max}$, $y_{min}$, $y_{max}$, Report);
                        continue := Report(t.p.right);
                ELSE RectangularRangeSearch := TRUE (" node fails test ")
                END
        ELSE RectangularRangeSearch := TRUE; (" empty tree ")
        END
END; (* of RectangularRangeSearch *)
```

5.2 *Circular Range Search*

For Circular Range Search, the Vincenty's inverse method is used to compute the geographical distance and azimuth between query point and others points.

```
Circular Range Search in modified PST
CircularRangeSearch( t : MPST , $q$, $r$)
VAR
    continue : BOOLEAN;
    q = query point;
    r = range(distance);
    $d_p$ = Vincenty's Inverse Method(q, t.p);
BEGIN
    IF t ≠ NIL , THEN
        IF ($d_p$, r), THEN
            continue := Report(t.p );
        ELSE IF t.p ≠ leaf , THEN
            CircularRangeSearch(t.p.left, q, r, Report);
            continue := Report(t.p .left);
_       ELSE IF t.p.left = leaf t.p.right ≠ NIL, THEN
            CircularRangeSearch(t.p.right, q, r, Report);
            continue := Report(t.p .right);
        ELSE CircularRangeSearch := TRUE (" node fails test ")
        END
    ELSE CircularRangeSearch := TRUE; (" empty tree ")
    END
END; (* of CircularRangeSearch *)
```

6 Selecting and Integrating Sub-images

Image reconstruction is being done in this system such that images taken by grabbed form Google earth. Image reconstruction allows creating a large field of view and can view the surrounding scene with real images. When the user is want to see the detail map of their interested place without download the entire Google earth or map there is needed to combine part of images from sub-images dataset because we have no ready make overview map for our research region. This system also contains the images integration to be panorama view with non-overlap region.

The sub-images of Yangon region are grabbed from Google earth and which are stored in data-set with their sub-image number, latitude, longitude. Generally, when selecting the sub-images, emphasizes on the image which must cover the user preference places, users' require services, user define range. We have used our own image data sets for experimenting. All necessary sub-images are automatically extracted from the data-set which is pre-constructed with modified PST. After sub-images are selected, these are integrated to be a panoramic view.

Table 2 Sample of Sub-Image Dataset

Image Number	Max Latitude	Min Longitude	Min Latitude	Max Longitude
8370	16 50 44.83	96 9 54.02	16 50 34.83	96 10 4.03
9712	16 48 34.70	96 8 14.02	16 48 24.89	96 8 23.95
10547	16 47 14.82	96 8 44.01	16 47 4.76	96 8 53.99
10552	16 47 14.85	96 9 34.00	16 47 4.78	96 9 43.99

This section describes image integration. Sub images are selected from data-set which is containing the sub-images of Yangon region. To obtain the sub-images, Yangon area is divided into 49680 parts on Google earth. Each of them is grabbed and which are stored in data-set with their sub-image number, latitude, longitude and position. These sub-images and detail information are constructed as modified PST as services. After sub-images are selected, these are integrated to be a panoramic view.

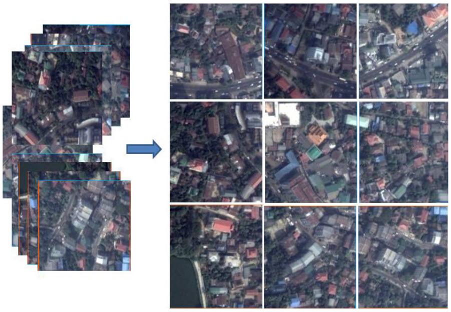


Fig. 4 Selecting Images form Database

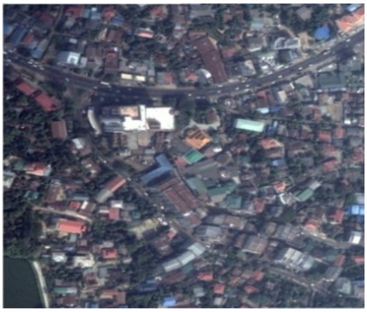


Fig. 5 The result of integrating the sub-images

7 Coordinate Transformation

For displaying the services on the map, there is needed to convert geographic coordinate to image coordinate of all services. Direct transformation by Grid-on-Grid method is used to rectify raw images which are built up by a rectangular array of pixels with variable values. Simplest transformation model is, (x',y') -> (x,y). Affine transformation allows for different scales in the directions of the two axes. Affine transformation is a linear transformation and relates two 2D cartesian coordinate systems through a rotation, a scale change in x- and y- direction and translations.

$$X' = s_x X \cos(\alpha) - s_y Y \sin(\alpha) + x_0$$

$$Y' = s_x X \sin(\alpha) + s_y Y \cos(\alpha) + y_0$$

where ;

α is angle of rotation.

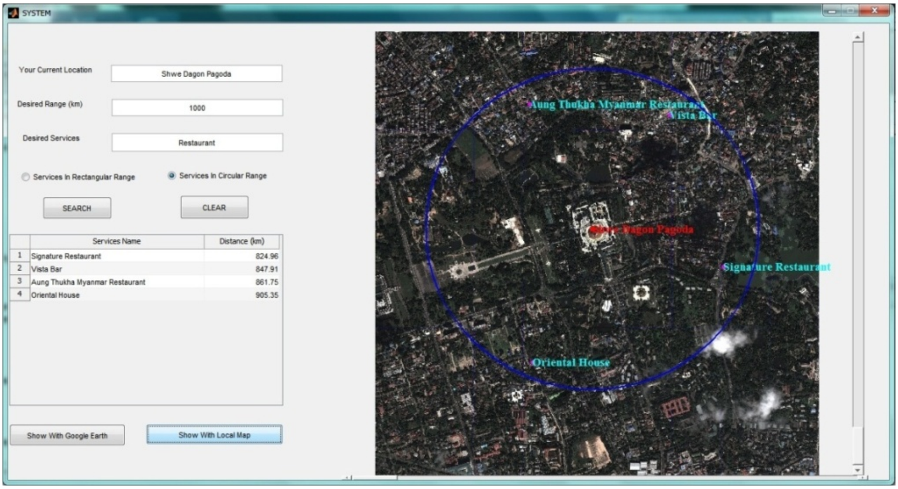
s_x and s_y are scale factor.

x_0 and y_0 are translati on parameters.

```
Coordinate Converting Algorithm
1. define α = 0;
2. define $x_0$ = origin shift in longitude at raster point (0,0);
3. define $y_0$ = origin shift in latitude at raster point(0,0) ;
4. calculate scale factor
   $latLim = [latMax \quad latMin],\ lonLim = [lonMax \quad lonMin]$
   $raster\ size = [m \quad n]$
   $s_x = \frac{lonMax - lonMin}{n},\ s_y = \frac{-latMax + latMin}{m}$
5. $p[]$ = latitudes to transform , $q[]$ = longitudes to transform
        $for\ i = 0 : p - 1;$
            $x = \frac{q[i] - x_0 + s_y y \sin(\alpha)}{s_x \cos(\alpha)}$
            $y = \frac{p[i] - y_0 + s_x x \sin(\alpha)}{s_y \cos(\alpha)}$
        $end$
```

8 Result and Discussion

In this paper, we demonstrate about the modifying priority search tree for range searching. In our system, we are only searching all nodes in defining range in dataset. This implies the time of *O(log k)* on depth of recursion and the number of nodes visited. Firstly, user needs to give their current location, desire distance range and desired services as input. After that user can choose rectangular range or circular range and the results can show on the map. These inputs are passed through as a query point ***q*** and range ***r*** to Range Searching Procedure in Modified PST.


SYSTEM
SEARCH
CLEAR
Show With Google Earth
Show With Local Map
Aung Thukha Myanmar Restaurant
Vista Bar
Shwe Dagon Pagoda
Signature Restaurant
Oriental House

Your Current Location: Shwe Dagon Pagoda

Desired Range (km): 1000

Desired Services: Restaurant

○ Services In Rectangular Range ● Services In Circular Range

	Services Name	Distance (km)
1	Signature Restaurant	824.96
2	Vista Bar	847.91
3	Aung Thukha Myanmar Restaurant	861.75
4	Oriental House	905.35

Fig. 6 Showing Services according to user preference

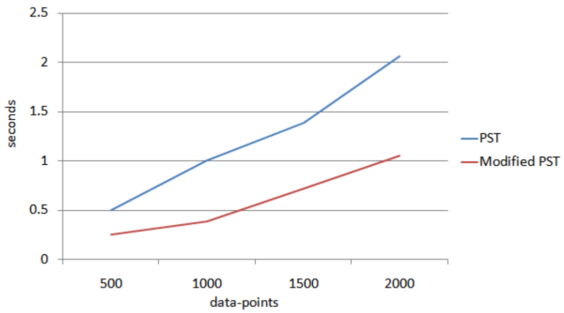


Fig. 7 Constructing time according to dataset

Table 3 Comparing Performance

Tree Types	Query Time			
	Rectangular Range		Circular Range	
Kd-Trees	Yes	$O(\sqrt{n} + k)$	No	-
Range Trees	Yes	$O(\log^2 n + k)$	No	-
Priority Search Trees	Yes	$O(\log n + k)$	No	-
Proposed System	Yes	$O(\log n + k)$	Yes	$O(\log n + k)$

9 Conclusion

The range searching system is useful for foreigner to easily search bus stop, restaurant, department station, etc.., and easily to estimate how far from there. Most current location based services are providing on online and need wireless connection. The system is based on geographic coordinates system and also useful in offline.

The system provide for searching desired services on desired distance range according to user preference.This system is planned to extend by enumerating the nearest neighbour for all services in the circular range. When the user is on the move, it is necessary to get the updated information in time, so we idealize to calculate the moving rang query and defying shortest part for end users.

References

1. McCreight, E.M.: Priority Search Tree. Society for Industrial and Applied Mathematics Journal **14**(2), May 1985
2. Hu, H., Lee, D.L.: Range Nearest-Neighbor Query. IEEE Transactions on Knowledge and Data Engineering **18**(1), January 2006. 24

3. Bentley, J.L., Friedman, J.H.: Data Structures for Range Searching. Computing Surveys **11**(4), December 1979
4. de Berg, M., Gudmundsson, J., Hammar, M., Overmars, M.: On R-trees with low query complexity. Computational Geometry **24**(3), 179–195 (2003)
5. De, M., Maheshwari, A., Nandy, S.C., Smid, M.: An in-place min-max priority search tree. In: 23rd Canadian Conference on Computational Geometry (2011)
6. Flatland, R.Y., Stewart, C.V.: Extending range queries and nearest neighbors. Computational Geometry **17** (2000)
7. Arya, S., Mount, D.M.: Approximate Range Searching. Computational Geometry: Theory and Applications **17**, 135–163 (2000)
8. http://en.wikipedia.org/wiki/Range_searching
9. http://www.movable-type.co.uk/scripts/latlong-vincenty-direct.html

A Novel Research Topic Ranking System in Academic Networks

Thi Thi Zin, Pyke Tin and Hiromitsu Hama

Abstract In today world, various types of communication networks such as academic, social, technological, business and etc. come into front. All of the networks are continuously growing and expanding in volume, velocity and variety like popular platform Big Data. Among them, the academic networks such as Alumni, Research Gate, Student Network, Teacher Network and so on provide a powerful abstraction of the academic structure and dynamics of diverse kinds of inter personal academic activities and interaction. Generally, the academic network contents such as research findings and educational concepts are created and consumed by the influences of all different academic navigation paths that lead to the challenging research issues. Therefore, identifying important and researcher relevant refined structures such as new research topics information or academic communities become major factors in modern decision making world.. In this paper, we propose a novel research topic ranking system in academic networks by using the research data relational graphs from academic media platform jointly with educational data to improve the relevance between research topics and researchers intentions (i.e., academic relevance). Specifically, we propose a stochastic model based Academic-Research Topic Ranking algorithm by taking academic value into account. By using some extensive experiments, we demonstrated the significant and effectiveness of the proposed academic-research topic ranking method.

Keywords Research topic rank · Academic networks · Stochastic model · Educational value · Academic similarity measures

---

T.T. Zin(✉)
Faculty of Engineering, University of Miyazaki, Miyazaki, Japan
e-mail: thithi@cc.miyazaki-u.ac.jp

P. Tin
Center for International Relations, University of Miyazaki, 1-1,
Gakuen kibanadai-nishi, Miyazaki 889-2192, Japan

H. Hama
Research Center for Industry Innovation, Osaka City University, 3-3-138, Sugimoto,
Sumiyoshi-ku, Osaka 558-8585, Japan

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_36

1 Introduction

Since the proposition of Journal Impact Factor [1] in 1963, the classical citation-based ranking scheme has been a standard criterion to rank journals and conferences. Recently, in Microsoft Academic Search, scholars of each domain are ranked by citation counts. Based on Microsoft's pioneer ranking system there have been proposed a lot of valuable different ranking methods, like h-index, g-index, citations per publication, etc [2-5]. However, we note that the reference of a paper cannot list all relevant publications and the citation relationships are not always available especially when related to copyright problem. Besides, we cannot evaluate a newly published paper before it is cited by others. However, the use of simple citation counts for researcher valuation may be flawed because it does not take into account that citation lists are not always complete or valid. While there are many new indices with interesting, unique attributes, none is a perfect metric for measuring scholars.

Therefore, in this paper, we propose a Multi-Quality Ranking system called MQ-Rank to rank a research topic in the field of academia or in the academic networks.

The main contribution of this paper is to present an application of stochastic methods based on Markov Chain model for retrieving meaningful academic information and ranking research topic information from academic commonly used educational platforms such as universities and various academic institutions.

The remainder of this paper is organized as follows. In Sections 2, we present some related works. The Section 3 introduces an overview of the proposed research topic ranking system in academic networks as described in Figure 1, and then followed by the technical details in section 4. We present and discuss the results of simulation experiments in section 5. Conclusions of this paper are presented in section 6.

2 Some Related Works

Our work in this paper is concerned with a stochastic modeling approach to research topic ranking system in academic network platforms. In this aspect, some authors [6-8] have used the popular Page Rank algorithm to retrieve research information from an academic learning dataset or from the World Learning networks. Since the Page Rank algorithm is based on the hyperlink structure of web graph, for such applications academic similarity among the research topics are used as academic links. It is worthwhile to mention that almost all applications of Page Rank and its extensions to a complete network and to any of its sub networks may give very different results [9-10]. Therefore a Page Rank based ranking algorithms applied to the entire Web can present very different results compared to the same algorithms applied to academic networks. To overcome this difficulty, we consider a first level of external links in addition to the academic network that is the set of resources from which the researchers enter in the academic networks.

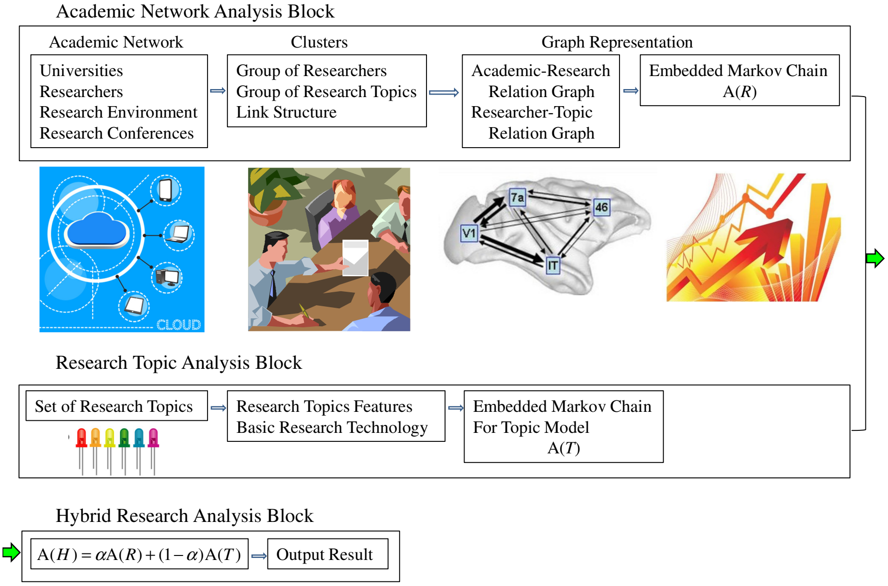

Academic Network Analysis Block
Academic Network
Universities
Researchers
Research Environment
Research Conferences
Clusters
Group of Researchers
Group of Research Topics
Link Structure
Graph Representation
Academic-Research Relation Graph
Researcher-Topic Relation Graph
Embedded Markov Chain A(R)
CLOUD
7a
46
V1
IT
Research Topic Analysis Block
Set of Research Topics
Research Topics Features
Basic Research Technology
Embedded Markov Chain For Topic Model A(T)
Hybrid Research Analysis Block
$A(H)=\alpha A(R)+(1-\alpha)A(T)$
Output Result

Fig. 1 Overview of Proposed Academic Network Ranking System

Moreover, much work has been done concerning with ranking research papers, journals and conferences in academic networks such as world learning networks [11-12]. However, the qualitative evaluation of any research topic ranking system in academic networks is not a trivial task since many qualitative measures, such as applicability, reliability, relevancy and attractiveness, are involved. Many previous studies [13-15] have shown that the reliability is strongly dependent on the academic interactions. Thus, academic research links for research topics and their relationships to the researchers play very significant roles in establishing research topic ranking system for social and academic network platforms.

3 The Proposed Research Topic Ranking System

In this section, we propose a general research topic ranking system in academic network platforms by using embedded Markov chains. Figure 1 shows the general framework of the proposed ranking system. The proposed ranking system contains three blocks namely; academic research network block, research content analysis block and hybrid research block. For each block an embedded Markov chain is established. First, we define the interest groups in the academic network block by using individuals and research topic of interests from an academic network. In the second two graphs; researchers-academic research topic relationship graph and research-community graph are constructed. Then the resultant graphs are combined to produce a research-academic interest group relationship graph. Finally, based on the academic-topic graph an embedded Markov chain is established. This Markov chain is denoted by A(R). We also construct an embedded Markov chain A(T) for the content-topic graph in the research content analysis block by using the co-occurrence matrix based on the co-occurring frequencies between the predefined research features of two topics. The two embedded Markov chains A(R) and A(T) are combined by using the following equation to produce academic research content Markov chain. This process is to be carried out in the hybrid ranking block.

$$A(H) = \alpha\ A(R) + (1-\alpha)\ A(T), \qquad 0 < \alpha < 1. \tag{1}$$

We then compute the stationary probabilities of the Markov chain A(H). These probabilities are used for the Multi-Quality Ranking system called MQ-Rank in academic networks.

4 Computational Procedures

In this section, we shall present some computational procedures for the Multi-Quality Ranking system called MQ-Rank in academic networks. In doing so we shall consider each block separately.

4.1 Academic Research Network Block

The main function of this block is to perform the derivation of the transition probability matrix A(R) for the research topic-academic relationship embedded Markov chain. This matrix is to be used in equation (1) to compute research topic rank as an academic relevance factor.

A. Research Interest Group rank:

For clarity, we first define a research interest group in an academic social network as a combination of set of researchers and set of interested research topics. Suppose there are *L groups* under consideration. They are denoted by $G_1, G_2, \ldots, G_L$.

Let R_i and T_i be the corresponding set of researchers and set of research topics in group G_i for $i = 1, 2, \ldots, L$. We then define a relationship measure between two groups in researcher's interests as described in equation (2).

$$\alpha_{ij} = \mathrm{Rel}(G_i, G_j) = \lambda\ I(R_i, R_j) + (1-\lambda)\ I(T_i, T_j) \qquad \text{for } 0 < \lambda < 1. \tag{2}$$

In this case, $I(R_i, R_j) = \#(R_i \cap R_j) \,/\, \#(R_i \cup R_j)$,

$I(T_i, T_j) = \#(T_i \cap T_j) \,/\, \#(T_i \cup T_j)$ where #(A) = the number of elements in set A.

After the pair-wised group relations are computed, we define the embedded Markov chain transition matrix for inter-group relationship in the interest of users as shown in equation (3).

$$A = [a_{ij}] \text{ where } a_{ij} = \alpha_{ij} \,/\, \Sigma\alpha_{ij} \tag{3}$$

Then we calculate the stationary distribution vector $\pi = [\pi_{1,}\pi_2, \ldots, \pi_L]$ by using the following equation.

$$\pi = A\pi \tag{4}$$

We will use the $\pi = [\pi_{1,}\pi_2, \ldots, \pi_L]$ as a group rank vector.

B. Group Correlation

In establishing an academic research topic ranking system for an academic network it is necessary to consider the group correlation as an academic influential factor. Thus, we need to define the group correlation between two groups G_i and G_j with respect to the interest of another membership community say G_k. Let $\rho_k(G_i, G_j)$ be the group correlation between G_i and G_j with respect to the interest of G_k. Then this correlation can be interpreted as the probability that a researcher in G_k will view the research topic of G_j after viewing the topics of G_i. Thus, we can also consider that if users in G_k are interested in images in G_i and G_i recommend G_j to G_k, then users in G_k may also be interested in images in G_j. Therefore, we can formulate the group correlation as

$$\rho_k(G_i, G_j) = (Rel(G_k, G_i) + RelG_k, G_j)) \quad Rel(G_i, G_j) \quad \alpha_i \ \alpha_j \tag{5}$$

where α_i and α_j *a*re group rank calculated in equation (4).

Group correlation defined in equation (5), we now present the procedure for computing rank with respect to academic research influence factor.

First, let us define the co-occurrence of two research topics T_l and T_m with respect to all academic groups described in equation (6).

$$d_{lm} = \Sigma_k \Sigma_j \Sigma_i \#\{(T_{l,} T_m) \mid T_l \in G_i \text{ and } T_m \in G_j\} \ \rho_k(G_i, G_j). \tag{6}$$

We can define the research-academic relationship embedded Markov chain transition matrix A(R) in equation (1) as follows.

$$A(R) = [p_{lm}], \tag{7}$$

where $$p_{lm} = d_{lm} / \Sigma d_{lm}.$$

From the Markov transition matrix in equation (7), we derive the stationary distribution vector and use the probabilities as the multi-quality ranks with respect to academic relationships.

In this block we shall compute the research topic rank with respect to research contents. In order to do so, we first decompose every 2 research groups into one.

From the sub-topic (segmented clusters) of each research topic in the research dataset, the content moments based feature vectors are extracted. The features of research in the database are pre-calculated and stored. To achieve better computational performance, the feature descriptors are calculated in off-line. Then the system computes the relationship measures among the segmented clusters. Then we rank the topics according to the best score and the most related topics are returned to achieve research topic rank with respect to research contents. Specifically for each cluster in the database information we find the feature vector f_j say.

We then compute the chi-square distance between f_i and f_j and used for relational measure between two topics. Let the related measures topic l and iopic m be v_{lm}.

We can have the content based embedded Markov transition matrix

$$\mathrm{A}(T) = [q_{lm}] \text{ where } q_{lm} = v_{lm} / \Sigma v_{lm} \, [p_{lm}]. \tag{9}$$

4.2 Hybrid Research Topic Rank Block

The function of this block is to compute research topic-content-academic rank or multi-quality rank described in this paper.. We first note that from academic network point of view, if interest group G_i which contains topic T_l is related to group *Gj* which contains topic Tm then the probability of researcher's jumping from topic T_l to topic Tm will be high. But from topic content point of view, a researcher may be attracted by some contents of topic T_l and then decide to study Tm which contains similar contents. Thus, we can conclude that both factors academic and information contents can play significant roles on establishing research topic ranking system in social and academic network platforms. Therefore, we define our hybrid research topic and researcher graph as a convex combination of topic content graph and social academic graph as described in equation (1).

We then substitute A(R) from equation (7) and A(T) from equation (9) into equation (1) to obtain A(H).

Let $\pi = [\pi_1, \pi_2, \dots, \pi_L, \dots]$ denotes the stationary distribution vector of A(H) then we have the following relation.

$$\pi \quad = \mathrm{A}(H)\,\pi. \tag{10}$$

We then rank the research topics according to ascending of the values in π. The highest value in π will give the highest rank of research topic with respect to academic value as well as research content important.

5 Experimental Simulation Results

As an illustration purpose, we consider here our own collected research topic data from *DBLP XML Records* available at *http://dblp.uni-trier.de/xml/* [16]. The DBLP dataset contains information about various research papers from various fields published over the years. This information includes the name of the research topics, and its sources. We built the relationship network defined as a graph, with each research topic representing a node and the relationship representing the edges in the graph, the edges being directed ones, directed from the researcher node to the topic node. We divide the collected data set into two communities or groups G_1 and G_2 (say) containing 20 topics in each group with 8 researchers in G_1 and 7 users in G_2.

$G_1 = \{R_1, T_1\}$ where R_1 stands for the set of researchers and T_1 stands for the set of topics.

$R_1 = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8\}$.

$T_1 = \{t_1, t_2, t_3, t_4, t_7, t_8, t_9, t_{10}, t_{13}, t_{14}, t_{15}, \mathrm{t}_{16}, t_{19}, t_{20}, t_{21}, t_{22}, t_{25}, t_{26}, t_{27}, t_{28}\}$.

$G_2 = \{R_2, R_2\}$ where R_2 stands for the set of researchers and G_2 stands for the set of topics.

$R_2 = \{r_3, r_4, r_5, r_6, r_8, r_9, r_{10}\}$.

$T_2 = \{t_3, t_4, t_5, t_6, t_9, t_{10}, t_{11}, t_{12}, t_{15}, t_{16}, t_{17}, t_{18}, t_{21}, t_{22}, t_{23}, t_{24}, t_{27}, t_{28}, t_{29}, t_{30}\}$.

By using equations (2), (3) and (4) in section 4, we obtain

A =

0.705882	0.294118
0.294118	0.705882

This gives us the community rank for G_1 and G_2 as $\pi_1=0.5$ and $\pi_2=0.5$.

Similarly by substituting these values in equation (5) we have the values of community correlation as follows.

$\rho_1(G_1, G_1) = 0.5$ $\rho_2(G_1, G_1) = 0.208333$

$\rho_1(G_1, G_2) = 0.147569$ $\rho_2(G_1, G_2) = 0.147569$

$\rho_1(G_2, G_1) = 0.147569$ $\rho_2(G_2, G_1) = 0.147569$

$\rho_1(G_2, G_2) = 0.208333$ $\rho_2(G_2, G_2) = 0.5$

Then by deriving the stationary distribution, we obtain the social image rank as follows.

Rank 1 images = $\{t_3, t_4, t_9, t_{10}, t_{15}, t_{16}, t_{21}, t_{22}, t_{27}, {}_{28}\}$,

Rank 2 images = $\{t_1, t_2, t_7, t_8, t_{13}, t_{14}, t_{19}, t_{20}, t_{25}, t_{26}\}$,

Rank 3 images = $\{t_5, t_6, t_{11}, t_{12}, t_{17}, t_{18}, t_{23}, t_{24}, t_{29}, {}_{30}\}$.

6 Conclusions

In this paper, we had proposed a novel framework of research topic ranking system in academic networks by using embedded Markov chain models. We have also explored how research contents and academic networks relevancy can be combined to understand researcher intentions. Illustrative computation procedures are also described by using samples from the data set collected from DBLP XML Records available at http://dblp.uni-trier.de/xml/. In future we shall work on by using more large data sets.

Acknowledgment. This work is partially supported by KAKENHI 25330133 Grant-in-Aid for Scientific Research(C).

References

1. Garfield, E., Sher, I.H.: New factors in the evaluation of scientific literature through citation indexing. American Documentation **14**(3), 195–201 (1963)
2. Weingart, P.: Impact of bibliometrics upon the science system: Inadvertent consequences? Scientometrics **62**, 117–131 (2005)
3. Zhang, S.D.: Judge a paper on its own merits, not its journal's. Nature **442**, 26 (2006)
4. Gigerenzer, G., Todd, P.M.: The ABC Research Group, Simple Heuristics That Make Us Smart. Oxford University Press, Oxford (1999)
5. Stringer, M.J., Sales-Pardo, M., Amaral, L.A.N.: Effectiveness of Journal Ranking Schemes as a Tool for Locating Information 3(2) (February 2008)
6. Ding, Y., Yan, E., Frazho, A., Caverlee, J.: PageRank for Ranking Authors in Co-citation Networks. Journal of the American Society for Information Science and Technology (2009)
7. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: Extraction and Mining of Academic Social Networks. In: Proc. of 14th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (2008)
8. Hirsch, J.E.: An Index to Quantify an Individual's Scientific Research Output. In: Proc. of National Academy of Sciences (2005)
9. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. World Wide Web Internet and Web Information Systems **54**(2), 1–17 (1998)
10. Bressan, M., Pretto, L.: Local computation of PageRank: the ranking side. In: ACM Conf. on Information and Knowledge Management, pp. 631–640 (2011)
11. Achievement Index Climbs the Ranks. Nature **448**(16), 737 (August 2007)
12. Ball, P.: Index aims for fair ranking of scientists. Nature **436**(7053), 900 (2005)
13. Egghe, L.: The influence of transformations on the h-index and the g-index. J. Am. Soc. Inform. Sci. Technol. **59**(8), 1304–12 (2008)
14. Serenko, A., Bontis, N.: Global ranking of knowledge management and intellectual capital academic journals. Journal of Knowledge Management **13**(1), 4–15 (2009)
15. Hama, H., Tin, P., Zin, T.T., Toriu, T.: A Stochastic Model for Popularity Measures in Web Dynamics. In: Proc. of 6th Intl. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, Germany, pp. 676–679 (October 15–17 2010)

SLA Guaranteed Migration Decision Model Using MCDM Approach

Zar Lwin Phyo

Abstract In Virtualized Data Center (VDC), the workload in each node changes dynamically overtime, so the workload in some machines may exceed the threshold, which not only cannot guarantee the quality of service requirements (QoS), but also waste resources. One of the key benefits of virtualization is reassignment of a virtual machine to another physical host can be done when resource shortage or poor utilization conditions occurs in host. However, in this case, efficient migration decision is required to reduce migration cost and ping-pong effect. In this paper, we propose SLA guaranteed migration decision model for resource management in VDC based on the integrated approach of Analytical Hierarchy Process (AHP) and the Preference Ranking Organization METHod for Enrichment Evaluations (PROMETHEE). In this case, AHP is used to assign weights to the criteria to be used in selection phase, while PROMETHEE is employed to determine the priorities of the alternatives.

Keywords AHP · PROMETHEE · Qos · SLA · Virtualized data center

1 Introduction

VDC contains physical and virtual servers which serve a variety of services including web services, file services etc. There are many applications these are running in VDC. Most of the applications are business critical applications with Quality-of-Service (QoS) requirements and these are running on virtual machines (VMs), and one or more VMs are mapped onto each physical machine (PM) in the system. In each VM, there are workload fluctuation from time to time and resources requirement changes overtime. An increase in workload can be handled by allocating more resources to it, if idle resources are available on the PM. If PM

Z.L. Phyo(✉)
Computer University, Meiktila, Myanmar
e-mail: {zarlwinphyo,cumtla}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_37

does not have enough or no idle resources to satisfy VM's requirement, the initial placement may lead to Service Level Agreement (SLA) violation and performance degradation.

In order to avoid SLA violation and performance degradation, live migration, the biggest advantage of virtualization, is used to handle. When VMs are needed to be migrated from one PM to another, the migration controller needs to decide which VMs should be migrated away and also decides where to migrate in datacenter [3]. When average utilization of a host reaches below defined threshold, underload condition, migration can also be made. In this case, all VMs from underloaded server are moved to other appropriate servers and then it is shut down in order to save energy cost. They can also be moved back into an active state and re-introduced into the running state when load peaks.

For a proper selection, the decision maker may need a large amount of data to be analyzed and many factors to be considered. The more criteria are considered, the more computational overhead it can take. But the insufficient number of criteria can cause inefficient decisions. Therefore, the decision maker is needed to balance the trade-off between number of criteria and the computational overhead. Therefore, this study utilizes a MCDM method (AHP) to determine the importance weights of evaluation criteria, and PROMETHEE to obtain the performance ratings of the feasible.

The rest of the paper is structured as follows: Section 2 presents the related work of this paper. Section 3 describes migration decision model and assumptions. Numerical evaluation is presented in section 4. It is followed by conclusion in section 5.

2 Related Work

In [9], the authors presented AppRAISE, a distributed management system for application performance control and dynamic resource allocation in virtualized server environments. They extended a traditional queuing model to represent application performance. They considered only CPU resource and only calculated the mean response time, which can be misleading. In [5], M.Tarighi presents a method to migrate VMs between cluster nodes using TOPSIS algorithm to find the most loaded server. In their system, they combine the TOPSIS algorithm and fuzzy theory. In their first level of implementation, they order the physical servers and decide which the most overloaded ones are. Then the second level of operation starts by ordering the virtual machines. After completing the ordering of two steps, then migration decisions are made to move the most overloaded virtual machines from the most overloaded physical machines to the least overloaded machine. But they did not consider the SLA in migration decision. Y.O.Yazir et al. [8], introduced a new approach for dynamic autonomous resource management in computing cloud. PROMETHEE II method was used for resource configuration. But they assumed all criteria have equal weigh in decision making. But considering the status of each parameter makes such an assumption unrealistic. For example, when the virtual machines are Web server, the probability of CPU saturation is more than RAM saturation. Because of the Web server are CPU-intensive load. Therefore, the weight of CPU and RAM influence are not equal. In this case, we use AHP method to calculate the weight of each criterion.

3 Migration Decision Model and Assumptions

We assume that the target virtualized data center offers multiple transactional Web services, and used share storage database. A set of heterogeneous servers are included in this system, each of which runs a Virtual Machine Monitor, such as VMWare or Xen. Physical resources of each server such as CPU, disks, memory, network are allocated among multiple virtual machines. Virtualization enables the flexible reduction and expansion of the resource capacity assigned to each VM. We made a few assumptions driven by the goal of meeting a reasonable trade-off between accuracy and practical solution time. We assume in target resource pool, there is enough capacity to create VMs for incoming requests and migrated VM can be accepted by target PM. Moreover, we assume that each virtual machine runs a single independent task. A server can be overloaded or underloaded in one or more dimensions such as CPU, Memory, Network, temperature and so on. In this model, we emphasize the overload or underload condition in CPU and Memory resources.

3.1 Virtual Resource Manager

There may occur an unbalanced condition in VDC when one host is overloading and another one is underloading. This situation can be handled by Virtual Resource Manager (VRM). VRM is centralized and is hosted on a dedicated server. VRM performs different operation upon underload condition and overload condition. In overload condition, VRM tries to move the most overloaded VM to other destination host. In this case, VRM ranks VMs of overload server by their score of descending order. The highest score of VM is most appropriate one to be migrated. And then, VRM finds the most preferable destination host for that VM. In this case, the PMs are also ranked in descending order. The machine with highest score is defined as the most overloaded server and the one with lowest score is the most appropriate destination host for VM to be migrated. In the case of underload condition, VRM ranks the hosts to know which one is the least loaded host and target host. Then VRM move all of the VMs on the least loaded machine to that appropriate target machines and then it is shutdown to save energy costs. In order to deal with this problem, we use the combination of AHP and PROMETHEE methods.

3.2 Migration Decision Based SLA Agreements

SLA is a document that includes a description about agreed service, service level parameters, guarantees, and actions and remedies for all cases of violations [1]. The SLA is very important as a contract between consumer and provider. The main idea of SLAs is to give a clear definition of the formal agreements about service terms like performance, availability and billing. Therefore SLA agreements should be considered when migration decision making. When overloading or underloading case occurs, VRM ranks the PMs according to their loads and suggests the PM that should be destination host according to their rank. However, the destination host must have enough capacity to their agree SLA value.

$$\mathbf{CAPPM(i) \geq SLAVM(i)} \qquad (1)$$

Availability capacity of destination host (PM CAPPM (i)) must greater than SLA agree value of migrated VM (SLAVM (i)) as shown in (1). Therefore, in this paper, we also consider agree SLA value in migration decision making process.

3.3 Analytical Hierichical Process(AHP)

AHP was developed by Saaty at 1980 [6]. It addresses how to determine the relative importance of a set of activities in a multi-criteria decision problem. The AHP method is based on three principles: first, structure of the model; second, comparative judgment of the alternatives and the criteria; third, synthesis of the priorities. In the literature [2], AHP has been widely used in solving many complicated decision-making problems.

3.4 PROMETHEE

The PROMETHEE (Preference Ranking Organization METhod for Enrichment Evaluation) method was developed by Brans and Vincke in 1985 [4]. In this paper, the PROMETHEE II method is employed to obtain the full ranking of the alternative criteria for VM migration decision. The procedural steps as involved in PROMETHEE II method are enlisted as below:

Step 1: Normalize the decision matrix using the following equation:

$$Rij = [X_{ij} - \min (X_{ij})]/[\max (X_{ij}) - \min (X_{ij})] \; (i = 1, 2, \ldots.., n \; ; j = 1,2,\ldots\ldots,m) \qquad (2)$$

where X_{ij} is the performance measure of i^{th} alternative with respect to j^{th} criterion.

Step 2: Calculate the evaluative differences of i^{th} alternative with respect to other alternatives.

Step 3: Calculate the preference function, $P_j(i,i')$

In PROMETHEE method, there are mainly six types of generalized preference functions. But these preference functions require the definition of some preferential parameters, such as the preference and indifference thresholds. However, in real time applications, it may be difficult to specify for the decision maker. To avoid this problem, a simplified preference function is adopted by V.M. Athawale and S. Chakraborty [7]. Therefore we used this preference function here:

$$P_j(i,i') = 0 \qquad \text{if } R_{ij} \le R_{i'j} \qquad (3)$$

$$P_j(i,i') = (R_{ij} - R_{i'j}) \qquad \text{if } R_{ij} > R_{i'} \qquad (4)$$

Step4: Calculate the aggregated preference function taking into account the criteria weights.

$$\pi(i,i') = \left[\sum_{j=1}^{m} w_j * P_j(i,i') \right] / \sum_{j=1}^{m} w_j \qquad (5)$$

where w_j is the relative importance (weight) of j^{th} criterion.

Step5: Determine the leaving and entering outranking flows for i^{th} alteernative as follows:

$$\varphi^{+}(i) = \frac{1}{n-1} \sum_{i'=1}^{n} \pi(i,i') \qquad (i \neq i') \qquad (6)$$

Entering (or negative) flow for i^{th} alternative,

$$\varphi^{-}(i) = \frac{1}{n-1} \sum_{i'=1}^{n} \pi(i',i) \qquad (i \neq i') \qquad (7)$$

where n is the number of alternatives.

Step 6: Calculate the net outranking flow for each alternative.

$$\varphi(i) = \varphi^{+}(i) - \varphi^{-}(i) \qquad (8)$$

Step 7: Determine the ranking of all the considered alternatives depending on the values of $\varphi(i)$. The higher value of $\varphi(i)$, the better is the alternative. Thus, the best alternative is the one having the highest $\varphi(i)$ value.

4 Numerical Evaluation

To make a numerical evaluation, we used the state information of resource pool of a VDC that is presented in [5]. As Table 1 indicates it has 5 PMs and 12 VMs.

Table 1 Physical Machines and its Virtual Machine

Physical Machine	**Virtual Machine**
PM_1	VM_1, VM_2
PM_2	VM_3
PM_3	VM_4, VM_5, VM_6, VM_7
PM_4	VM_8, VM_9, VM_{10},
PM_5	VM_{11}, VM_{12}

Table 2 shows the state information of Physical Machine. In this table, the CPU usage of each node in percent (CPU%), RAM usage of each node in percent (RAM %), Network usage in percent (NET%), RAM Capacity of each node, Network Bandwidth (Network BW) and CPU Clock Speed in GHz of each node are presented. According to this table, physical machine 3 has obtained the highest score. It means that this node is in danger and migration is inevitable. Therefore, virtual machines of PM3 are needed to know which one should be migrated away.

Table 2 State Information of PMs

PM	CPU%	RAM%	NET%	CPU Clock Speed	RAM Capacity	Network BW
PM1	15	32	13	2	2	100
PM2	0	2	0	1.2	1	1000
PM3	81	60	40	1.8	2	100
PM4	70	49	72	3.2	1	1000
PM5	53	70	16	2.4	6	100

4.1 VM Selection

CPU usage percentage is the most preferable criteria for the data center's administrator because of the most of the servers are running CPU intensive application in this resource pool.

Calculate the Weight of Criteria for VM Selection. In this phase, the weights of the criteria to be used in evaluation process are calculated by using AHP method. In this phase, the experts in the decision making team are given the task of forming individual pairwise comparison matrix by using the scale given in Table 3.

Table 3 Nine-Point Intensity of Importance Scale and its Description

Definition	Intensity of Importance
Equally Importance	1
Moderately more Importance	3
Strongly more Importance	5
Very Strongly more Importance	7
Extremely more Importance	9
Intermediate values	2 , 4 , 6 , 8

Table 4 State Information of Virtual Machines

	CPU%	RAM%	NET%	RAM Usage
VM_1	15	23	13	0.6
VM_2	0	0	0	0
VM_3	60	67	58	0.4
VM_4	54	56	72	1

Table 5 Pairwise Comparison Matrix

	C_1	C_2	C_3	C_4
C_1	1	3	5	2
C_2	0.33	1	2	0.5
C_3	0.2	0.5	1	0.5
C_4	0.5	2	2	1

For virtual machine migration, four criteria (C) were considered and their state information is shown in Table 4. In this case, C_1, C_2 andC_3 are defined as benefit criteria. C_4 is defined by cost criteria because virtual machine which has the lowest RAM usage is the optimal virtual machine that should be migrated. The more memory pages are used by the VM, the more migration time it will take. The pairwise comparison matrix that represents the relative importance of one criterion over another that is defined by decision maker is presented in Table 5. According to Table 6, CPU% and RAM usage are determined as the two most important criteria in VM selection process by AHP. Consistency ratio of the pairwise comparison matrix is calculated as $0.005 < 0.1$. So the weights are shown to be consistent and they are used in the selection process. According to Table 7, VM_3 should be migrated away.

Table 6 Weight Values of each Criteria

Criteria	Weight Values
C_1	0.489
C_2	0.162
C_3	0.098
C_4	0.251

Table 7 Net Outranking values for Virtual Machines

	Net Flow	Rank
VM_1	-0.234	3
VM_2	-0.286	4
VM_3	0.364	1
VM_4	0.154	2

4.2 Target Server Selection

For PM selection, the pairwise comparison matrix that represents the relative importance of one criterion over another that is defined by decision maker and weight value of each criterion are shown in Table 8 and Table 9.

Table 8 Pairwise Comparison Matrix

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	1	3	5	1	3	5
C_2	0.33	1	2	0.33	1	2
C_3	0.2	0.5	1	0.2	0.5	1
C_4	1	3	5	1	3	5
C_5	0.33	1	2	0.33	1	2
C_6	0.2	0.5	1	0.2	0.5	1

Table 9 Weight Values of each Criteria

Criteria	Weight Values
C_1	0.289
C_2	0.120
C_3	0.064
C_4	0.341
C_5	0.120
C_6	0.064

According to Table 9,CPU % and CPU Clock Speed are also determined as the two most important criteria in the PM selection process by AHP. Consistency ratio is calculated as $0.021 < 0.1$. So the weights are shown to be consistent and they are used in the selection process.The final result is shown in Table 3. According to this table PM 3 is the most overloaded server and PM 2 is the least loaded server, when PROMETHEE is used for the target server selection according to the collected data. Therefore, in this case PM 2 is chosen as the destined server.

Table 10 Net Outranking Values for Physical Machines

	Net Flow	Rank
PM1	0.381	3
PM2	-0.052	5
PM3	0.731	1
PM4	0.383	2
PM5	0.0.89	4

4.3 SLA Evaluation

According to the final result of Table 10, PM 2 is chosen as the destined server. However we need to check SLA agreement is violated or not in that PM 2 by using (1). If violation can occur in PM 2, we must choose other host although PM 2 is suitable host according to ranking table. Otherwise, overloaded VM is migrated to PM 2.

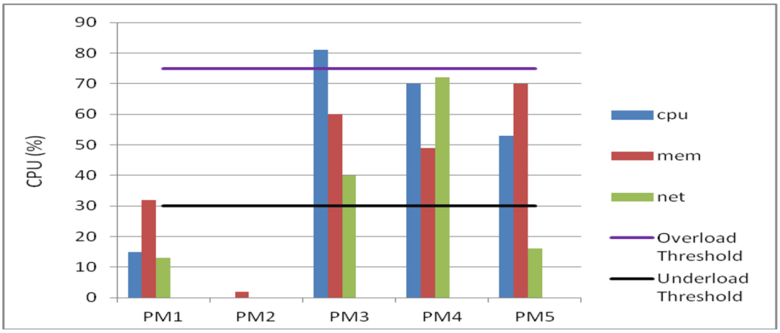


Fig. 1 Overloading occurs at PM3

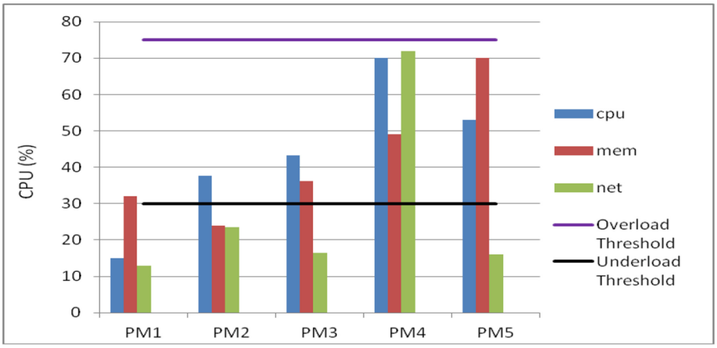


Fig. 2 After Migration Workloads of Overloaded Machine

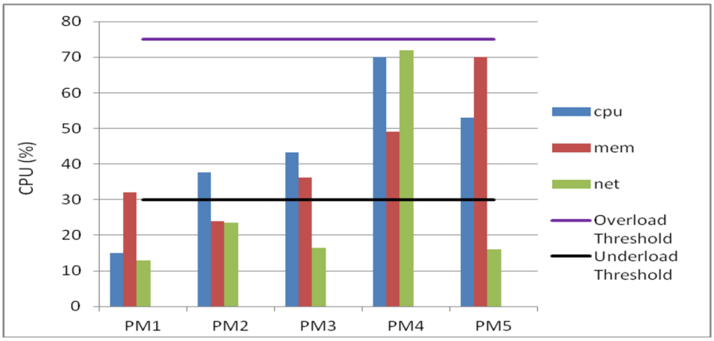


Fig. 3 Underloading occurs at PM1

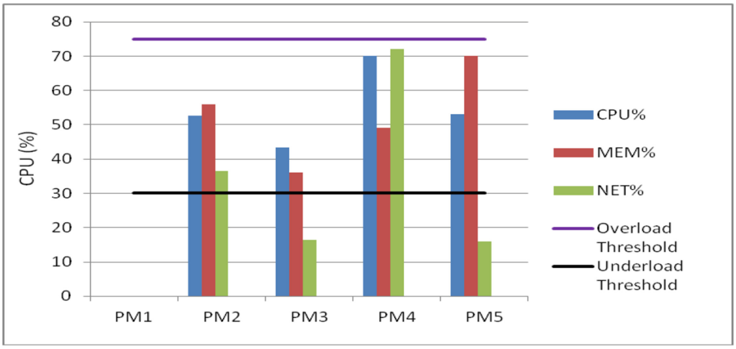


Fig. 4 After Migrating all Workloads of Underloaded Machine

5 Conclusion

In the case of resource shortage or poor utilization of host, reassignment of a virtual machine to another physical host is required in VDC. To migrate, we must know which virtual machine needs to be migrated and which host must be destined. To solve this problem, SLA guaranteed migration decision model is proposed. In this case, AHP is used to assign weights to the criteria to be used in selection phase, while PROMETHEE is employed to determine the priorities of the alternatives. The weights obtained from AHP are included in decision making process by using them in PROMETHEE computations and the alternative priorities are determined based on these weights. Then we also evaluate whether the decision meets the corresponding SLA or not.

References

1. Andrieux, A., et al.: Web services agreement specification (WSAgreement) (2004)
2. Chan, F.T.S. et al.: Global Supplier Selection: a fuzzy-AHP approach. International Journal of Production Research doi:10.1080/00207540600787200
3. Mevag, I.: Towards Automatic Management and Live Migration of Virtual Machines. Thesis of Oslo University (May 2007)
4. Doumpos, M., Zopoundidis, C.: A multi-criteria classification approach based on pair-wise comparison. European Journal of Operation Research **158**, 378–389 (2004)
5. Tarighi, M., et.al.: A New Model for Virtual Machine Migration in Virtualized Cluster Server Based on Fuzzy Decision Making. Journal of Telecommunications **1**(1) (February 2010)
6. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)

7. Athawale, V.M., et.al.: Facility Location Selection using PROMETHEE II Method. In: Proceedings of the International Conference on Industrial Engineering and Operations Management Dhaka, Bangladesh (2010)
8. Yazir, Y.O., et.al.: Dynmic Resource allocation in Computing Clouds using Distributed Multiple Criteria Decision Analysis. In: IEEE 3rd International Conference on Cloud Computing (2010)
9. Wang, Z., et.al.: AppRAISE: Application-Level Performance Management in Virtualized Server Enviruonments. IEEE Transactions on Network and Service Management 6(4) (December 2009)

Part V

Recent Progress in Computational Electromagnetic Dynamics

Evaluation of Localized Polarized Light Generated by Nano Antennas for High-Density All-Optical Magnetic Recording

**Shinichiro Ohnuki, Keisuke Tatsuzawa, Yuta Takano,
Yoshito Ashizawa and Katsuji Nakagawa**

Abstract A next generation technology of magnetic recording is proposed and verified in terms of electromagnetic simulation. Our method is based on all-optical magnetic recording using nano antennas together with particulate media. This novel method can achieve ultra-fast and high-density magnetic recording: the recording speed is about 100,000 times faster than that for conventional technology and the recording density becomes over a few Tbit/inch^2. We will evaluate the localized light generated by various nano antennas and discuss the stability of our recording method.

Keywords All-optical magnetic recording · Nano antennas · Electromagnetic simulation

1 Introduction

A next generation technology of magnetic recording is proposed and verified in terms of electromagnetic simulation. Our method is based on all-optical magnetic recording with circularly polarized light and it can be considered as one of the promising technologies for ultra-fast magnetic recording [1]: the recording speed can be 100,000 times faster than that for conventional method.

We have designed nano antennas and particulate recording media for realizing high-density all-optical recording [2-5]. Using our method, circularly polarized light is successfully localized in nano scale and one-bit information is stored

---

S. Ohnuki(✉) · K. Tatsuzawa · Y. Takano · Y. Ashizawa · K. Nakagawa
College of Science and Technology, Nihon University, 1-8-14 Surugadai,
Kanda, Chiyoda-ku, Tokyo 101-8308, Japan
e-mail: ohnuki.shinichiro@nihon-u.ac.jp

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_38

inside each particulate recording medium, i.e. the recording density becomes over a few Tbit/inch^2. Recently, such designed nano antennas have been fabricated due to the progress of nano technology. However, the shapes are slightly different from the numerically designed one.

In this paper, we will discuss and verify the stability of our proposed method for realistic antenna models whose surfaces are designed considering nano fabrication. The generation condition of circularly polarized light will be clarified.

2 Computational Methods

To design and verify our magnetic recording method, we perform time-domain electromagnetic simulation using the finite-difference time-domain (FDTD) method. The motion equation of electrons is considered as the auxiliary differential equation (ADE) to estimate frequency dispersion of nano antennas and particulate recording media [6]. The finite-difference forms of the magnetic and electric fields, the current, and the polarization can be expressed as follows:

$$\mathbf{H}^{n+\frac{1}{2}} = \mathbf{H}^{n-\frac{1}{2}} - \frac{\Delta t}{\mu_0}\left(\nabla \times \mathbf{E}^n\right), \tag{1}$$

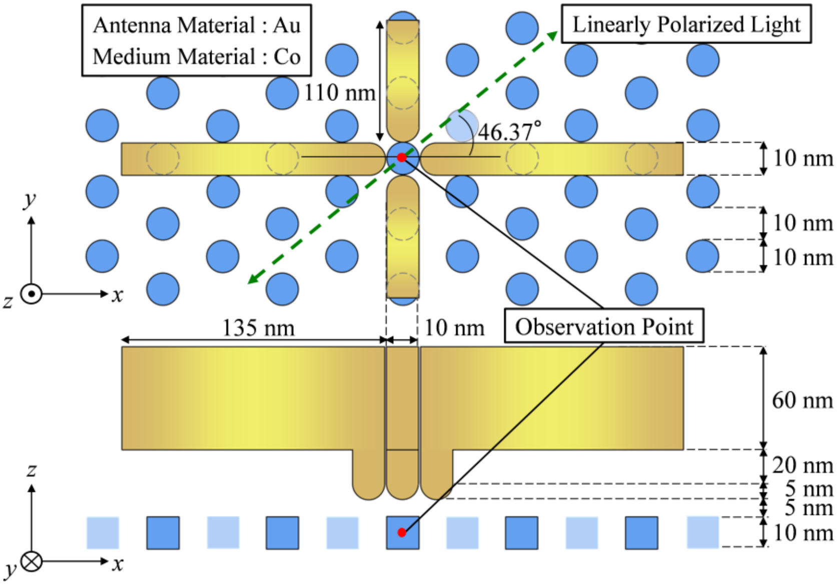

Antenna Material : Au
Medium Material : Co
Linearly Polarized Light
110 nm
46.37°
10 nm
10 nm
10 nm
y
z
x
135 nm
10 nm
Observation Point
60 nm
20 nm
5 nm
5 nm
10 nm
z
y
x

Fig. 1 Computational model of ultra-fast and high-density magnetic recording system.

$$\begin{aligned}\mathbf{E}^{n+1} &= C_1\mathbf{E}^n \\ &+ C_2\left[\nabla\times\mathbf{H}^{n+\frac{1}{2}} - \frac{1}{2}\sum_{l=0}^{K}\left\{(1+\alpha_l)\mathbf{J}_l^n - \gamma_l\mathbf{P}_l^n\right\}\right],\end{aligned} \tag{2}$$

$$\mathbf{J}_l^{n+1} = \alpha_l\mathbf{J}_l^n + \beta_l\left(\mathbf{E}^{n+1} + \mathbf{E}^n\right) - \gamma_l\mathbf{P}_l^n\ , \tag{3}$$

$$\mathbf{P}_l^{n+1} = \mathbf{P}_l^n + \frac{\Delta t}{2}\left(\mathbf{J}_l^{n+1} + \mathbf{J}_l^n\right), \tag{4}$$

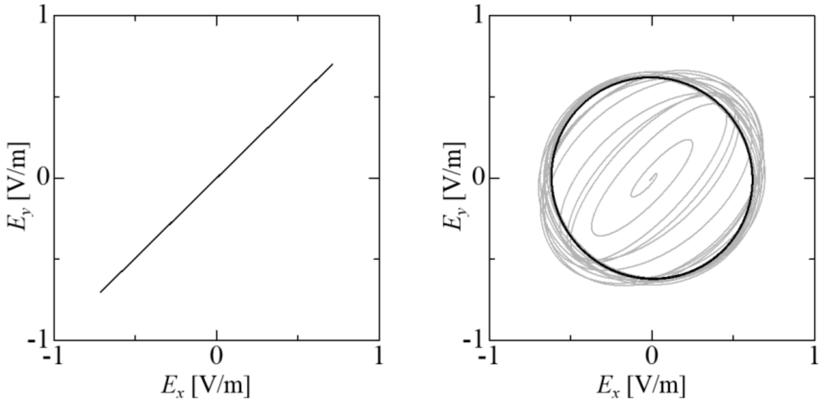
where

$$\begin{cases} C_1 = \dfrac{2\varepsilon_0 - \Delta t\sum_{l=0}^{K}\beta_l}{2\varepsilon_0 + \Delta t\sum_{l=0}^{K}\beta_l}, \quad C_2 = \dfrac{2\Delta t}{2\varepsilon_0 + \Delta t\sum_{l=0}^{K}\beta_l}, \\ \alpha_l = \dfrac{1}{\xi_l}\left\{1 - \dfrac{\Delta t}{4}\left(2\nu_l + \omega_l^2\Delta t\right)\right\}, \\ \beta_l = \dfrac{\varepsilon_0 A_l \omega_p^2 \Delta t}{2\xi_l}, \quad \gamma_l = \dfrac{\omega_l^2\Delta t}{\xi_l}, \\ \xi_l = 1 + \dfrac{\Delta t}{4}\left(2\nu_l + \omega_l^2\Delta t\right). \end{cases} \tag{5}$$

3 Numerical Results

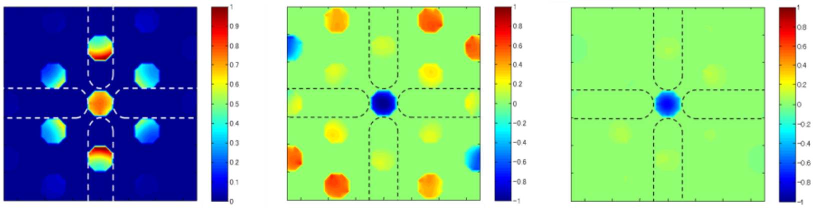
Figure 1 shows the designed nano antenna with the particulate recording media. The material of the antenna is gold and the recording media is cobalt. The dispersion characteristics are assumed to be the Lorentz-Drude model [7]. The size and placement of the particulate media are selected to realize over 2 Tbit/inch^2 recording density. The observation point is at the center of the target particle which is placed just below the antenna gap. The incident light impinges on the nano antenna and propagates in the negative *z*-direction.

The polarization of the incident light is linear as the Lissajous plot of the electric fields in Figure 2(a). Since the nano-antennas is designed to make the 90° phase difference of plasmon resonances between E_x and E_y components, the polarization at the observation point becomes circular after 40 fs as shown in Figure 2(b).



(a) (b)

Fig. 2 Lissajous curves. (a) incident light and (b) localized circularly polarized light inside the target media.



(a) (b) (c)

Fig. 3 Field distribution of modified Stokes parameters. (a) intensity, (b) circularity, and (c) figure of the merit.

The field distribution of modified Stokes parameters [3,4] are evaluated. From Figure 3, the enhanced and localized circularly polarized light can be observed only inside the target medium. This is the ideal condition of achieving higher-density all-optical magnetic recording.

The surface of the antenna is slightly modified to consider the margin of nano fabrication. Figure 4 shows a realistic image of nano antennas. The dashed line indicates the original flat surface of the numerically designed one. The margin is selected to consider our FIB equipment. Figure 5(a) shows the Lissajous curve for the modified antenna. Instead of the circularly polarized light, the elliptically polarized light is generated at the observation point. Since the plasmon mode is strongly affected to the shape of the antenna, the phase difference between E_x and E_y does not hold 90° anymore. If we can reshape the antenna and make exactly the same as the numerically designed one, the localized circularly polarized light should be obtained, however, the current nano fabrication technique does not allow us to do so easily. Therefore, we introduce another remedy. The nano antenna in Figure 1 was designed to obtain the circularly polarized light for the linearly polarized incidence. We will control the polarization of the incident light for the realistic antenna models. Figure 5(b) shows the Lissajous curve for the specifically designed incident light. The localized circularly polarized light can be obtained.

The last example is a more general case: the antenna surface varies randomly within the margin of nano fabrication. We still can generate the circularly polarized light with the 90° phase difference even for this special structure, if the incident light is carefully designed.

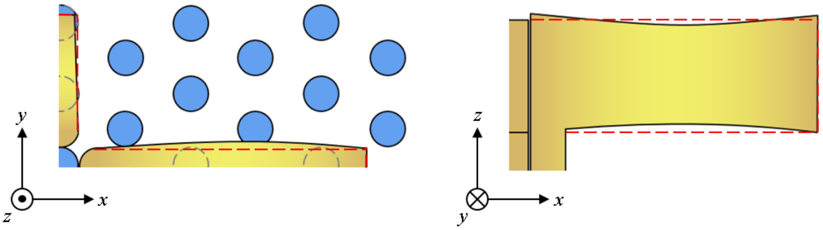
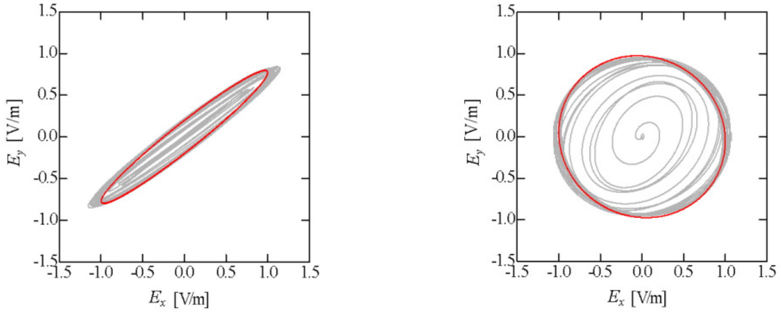


Fig. 4 Realistic image of a nano antenna.



(a) (b)

Fig. 5 Lissajous curves for different incident lights. (a) linear polarization, (b) Elliptical polarization.

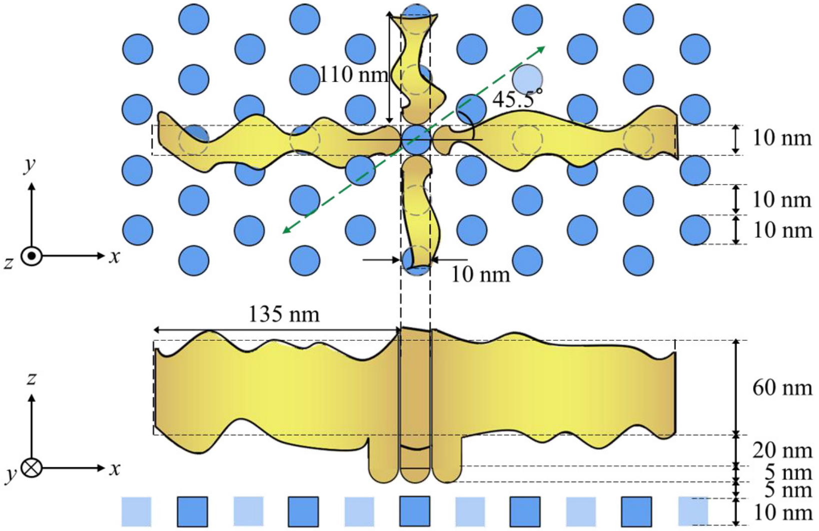


Fig. 6 Realistic antenna model: random surface case.

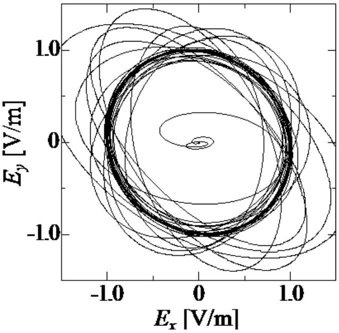


Fig. 7 Lissajous curves.

4 Conclusions

A next generation technology of magnetic recording is verified in terms of time-domain electromagnetic simulation. Using our designed method, nano antennas and particulate recording media, circularly polarized light can be successfully localized in nano scale and high-density all-optical magnetic recording can be realized. The generation condition of circularly polarized has been clarified for realistic antenna models whose surfaces were designed considering nano fabrication.

Acknowledgements. This work was partly supported by Grant-in-Aid for Scientific Research (C) (26420321) and MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2013-2017.

References

1. Stanciu, C.D., Hansteen, F., Kimel, A.V., Kirilyuk, A., Tsukamoto, A., Itoh, A., Rasing, T.: All-Optical Magnetic Recording with Circularly Polarized Light. Phys. Rev. Lett. 99, 047601-1–047601-4 (2007)
2. Nakagawa, K., Ashizawa, Y. Ohnuki, S., Itoh, A., Tsukamoto, A.: Confined Circularly Polarized Light Generated by Nano-Size Aperture for High Density All-Optical Magnetic Recording. J. Appl. Phys. **109**(7), 07B735 (2011)
3. Ohnuki, S., Kato, T., Takano, Y., Ashizawa, Y., Nakagawa, K.: Design and Numerical Verification of Plasmonic Cross Antennas to Generate Localized Circularly Polarized Light for All-Optical Magnetic Recording. Radio Science **50**, 29–40 (2015)
4. Kishimoto, S., Okada, T., Ohnuki, S., Ashizawa, Y., Nakagawa, K.: Efficient Analysis of Electromagnetic Fields for Designing Nanoscale Antennas by Using a Boundary Integral Equation Method with Fast Inverse Laplace Transform. Progress in Electromagnetics Research **146**, 155–165 (2014)
5. Tamura, K., Ota, T., Ashizawa, Y., Tsukamoto, A., Itoh, A., Ohnuki, S., Nakagawa, K.: Circularly Polarized Light Generated by Plasmon Antenna for All-Optical Magnetic Recording. Journal of the Magnetics Society of Japan **37**(3–2), 115–118 (2013)
6. Yamaguchi, T., Hinata, T.: Optical near-field analysis of spherical metals: Application of the FDTD method combined with the ADE method. Opt. Express **15**(18), 11481–11491 (2007)
7. Rakić, A.D., Djurišić, A.B., Elazar, J.M., Majewski, M.L.: Optical Properties of Metallic Films for Vertical-Cavity Optoelectronic devices. Appl. Opt. **37**, 5271–5283 (1998)

A Formulation of Electromagnetic Scattering by Defected Periodic Cylinder Array Embedded in Dielectric Slab

Koki Watanabe

Abstract This paper proposes a spectral-domain formulation of the electromagnetic scattering from a defected periodic cylinder array embedded in dielectric slab. The present formulation is based on the pseudo-periodic Fourier transform (PPFT), which makes it possible to apply the conventional grating formulations to the problems of imperfectly periodic structures. The formulation is also based on the recursive transition-matrix algorithm with the lattice sums technique, which is widely used for the problems of periodic cylinder array.

Keywords Defected periodic cylinder array · Pseudo-periodic fourier transform · Recursive transition-matrix algorithm

1 Introduction

The periodic structures are widely used for optical, photonics, and electronics applications. Furthermore, small collapses of the periodicity are sometimes useful to control the electromagnetic wave propagation in the periodic structures. When the incident field is not the plane-wave or the structure is not perfectly periodic, the fields generally have continuous spectra in the wavenumber space, and an artificial discretization in the wavenumber space is necessary for the spectral-domain approach. Watanabe and Yasumoto introduced the pseudo-periodic Fourier transform (PPFT) [1] to consider the discretization scheme for periodic structures for non-plane incident waves. Since the PPFT makes the fields pseudo-periodic, the conventional grating theories with the use of the generalized Fourier series expansions [2] also

---

K. Watanabe(✉)
Fukuoka Institute of Technology, 3-30-1 Wajirohigashi,
Higashi-ku, Fukuoka 811-0295, Japan
e-mail: koki@fit.ac.jp

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_39

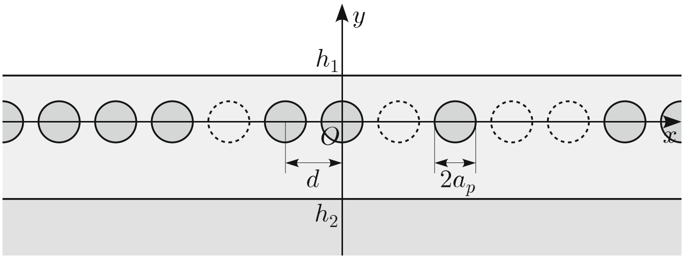


Fig. 1 Geometry under consideration.

become possible to be applied to the scattering problems of imperfectly periodic structures.

This paper deals with the electromagnetic scattering from a defected periodic cylinder array embedded in dielectric slab and propose a spectral-domain formulation, in which the PPFT concept combined with the recursive transition-matrix algorithm (RTMA) [3, 4] with the lattice sums technique [5]. The RTMA is one of the most commonly used approaches to the electromagnetic scattering from a finite number of parallel cylinders, and the lattice sums technique extended it to the plane-wave scattering from periodic cylinder array.

2 Settings of the Problem

The present paper considers the scattering problem of electromagnetic fields with a time-dependence $\exp(-i\,\omega\,t)$ from a defected periodic circular cylinder array embedded in a dielectric slab. Figure 1 shows an example of the structures under consideration. The structure consists of three homogeneous and isotropic layers, in which the permittivity and the permeability are denoted by (ε_1, μ_1) for the upper region $y > h_1$, (ε_2, μ_2) for the slab region $h_2 < y < h_1$, and (ε_3, μ_3) for the lower region $y < h_2$. The slab region includes circular cylinders, which are infinitely long and situated parallel to the z-axis. The cylinders are periodically located in the slab region, but some cylinders are removed. The original periodic array is formed by identical cylinders with the permittivity ε_p, the permeability μ_p, and the radius a_p, and the period is denoted by d. We suppose that the axis of the qth-cylinder in the original array is located at $(x, y) = (q\,d, 0)$ for a positive constant d. Then, to indicate the removed cylinders, we introduce a notation $\mathcal{D}$, which is finite subset of the integer set $\mathbb{Z}$. If $m \in \mathcal{D}$, the mth-cylinder is removed. The parameters are chosen so as that these cylinders are not overlapped, and the wavenumber in each medium is denoted by $k_r = \omega\sqrt{\varepsilon_r\,\mu_r}$ for $r = 1, 2, 3, p$. The electromagnetic fields are supposed to be uniform in the z-direction, and two-dimensional scattering problem is here considered. Two fundamental polarizations are expressed by TM and TE, in which the magnetic and the electric fields are respectively perpendicular to the z-axis. We denote the z-component of the electric field for the TM-polarization and the z-component of the magnetic field for the TE-polarization by $\psi(x, y)$, and show the formulation for both polarizations simultaneously. The incident field is supposed to illuminate from outside the slab region.

3 Tools for the Formulation

3.1 Pseudo-periodic Fourier Transform

Let $f(x)$ be a function of x and d be a positive real constant. Then the PPFT is defined by:

$$\overline{f}(x; \xi) = \sum_{m \in \mathbb{Z}} f(x - m\,d)\, e^{i\,m\,d\,\xi} \tag{1}$$

which is implicitly assumed to converge. This transform uses a transform parameter ξ, and the inverse transform is formally given by integrating on ξ as

$$f(x) = \frac{1}{k_d} \int_{-k_d/2}^{k_d/2} \overline{f}(x; \xi)\, \mathrm{d}\xi \tag{2}$$

where $k_d = 2\,\pi/d$. It is worth noting that the integration is over a finite interval.

The transformed function $\overline{f}(x; \xi)$ has a pseudo-periodic property with the pseudo-period d in terms of x: $\overline{f}(x - m\,d; \xi) = \overline{f}(x; \xi)\, e^{-i\,m\,d\,\xi}$ for any integer m, and can be therefore expressed in the generalized Fourier series expansion [2] as

$$\overline{f}(x; \xi) = \sum_{n \in \mathbb{Z}} \overline{f}_n(\xi)\, e^{i\,\alpha_n(\xi)\,x} \tag{3}$$

with $\alpha_n(\xi) = \xi + n\,k_d$. The coefficients $\overline{f}_n(\xi)$ are given by the orthogonal relation as

$$\overline{f}_n(\xi) = \frac{1}{d} \int_{-d/2}^{d/2} \overline{f}(x; \xi)\, e^{-i\,\alpha_n(\xi)\,x}\, \mathrm{d}x. \tag{4}$$

Using the definition of PPFT given by Eq. (1), the coefficients are also expressed as

$$\overline{f}_n(\xi) = \frac{1}{d} \int_{-\infty}^{\infty} f(x)\, e^{-i\,\alpha_n(\xi)\,x}\, \mathrm{d}x. \tag{5}$$

The expression in Eq. (5) implies that the generalized Fourier coefficient $\{\overline{f}_n(\xi)\}$ is equivalent to the Fourier integral of $f(x)$ with the wavenumber $\alpha_n(\xi)$.

Also $\overline{f}(x; \xi)$ has a periodic property with the period k_d in terms of ξ: $\overline{f}(x; \xi - m\,k_d) = \overline{f}(x; \xi)$ for any integer m. Therefore, the Brillouin zone becomes the analysis region in terms of ξ. If we limit the analysis region to $-k_d/2 < \xi \le k_d/2$, the nth-order coefficient $\overline{f}_n(\xi)$ gives the Fourier integral in the nth-order Brillouin zone. In other words, the PPFT folds the Fourier integral into the 0th-order Brillouin zone.

3.2 Expansion Bases

The PPFT field in the homogeneous regions can be expressed in the plane-wave expansion [1]. The plane-wave expansion bases are denoted by column matrices $\boldsymbol{f}_r^{(\pm)}(x, y; \xi)$ whose nth-components are given by

$$\left(\boldsymbol{f}_r^{(\pm)}(x, y; \xi)\right)_n = e^{i\,(\alpha_n(\xi)\,x \pm \beta_{r,n}(\xi)\,y)} \tag{6}$$

$$\alpha_n(\xi) = \xi + n\,k_d, \qquad \beta_{r,n}(\xi) = \sqrt{{k_r}^2 - \alpha_n(\xi)^2} \tag{7}$$

for $r = 1, 2, 3$. The present formulation will use also the cylindrical-wave expansions to express the fields inside the slab region but outside the cylinders. To express the expansion bases, we use column matrices $\boldsymbol{g}^{(Z)}(x, y)$ whose nth-components are given by

$$\left(\boldsymbol{g}^{(Z)}(x, y)\right)_n = Z_n(k_2\,\rho(x, y))\,e^{i\,n\,\phi(x,y)} \tag{8}$$

$$\rho(x, y) = \sqrt{x^2 + y^2}, \qquad \phi(x, y) = \arg(x + i\,y), \tag{9}$$

where Z specifies the cylinder functions associating to the cylindrical-wave bases in such a way that $Z = J$ denotes the Bessel function and $Z = H^{(1)}$ denotes the Hankel function of the first kind.

The translation relations for converting the reference points of the plane-wave expansion bases are given by

$$\boldsymbol{f}_r^{(\pm)}(x, y - y_p; \xi) = \boldsymbol{F}_r(\pm\left(y_q - y_p\right); \xi)\,\boldsymbol{f}_r^{(\pm)}(x, y - y_q; \xi) \tag{10}$$

$$(\boldsymbol{F}_r(y; \xi))_{n,m} = \delta_{n,m}\,e^{i\,\beta_{r,n}(\xi)\,y}. \tag{11}$$

Graf's addition theorem yields the transform relations for converting the reference points of the bases. Let (x_p, y_p) and (x_q, y_q) be the reference points of the cylindrical-wave basis. Then, when (x, y) is inside a circle with center (x_q, y_q) and radius $\rho(x_p - x_q, y_p - y_q)$, the transform relations are given by

$$\boldsymbol{g}^{(Z)}(x - x_p, y - y_p) = \boldsymbol{G}^{(Z)}(x_q - x_p, y_q - y_p)\,\boldsymbol{g}^{(J)}(x - x_q, y - y_q). \tag{12}$$

$\boldsymbol{G}^{(Z)}(x, y)$ denotes the Toeplitz matrix whose (n, m)-components are given by $(n - m)$th-component of $\boldsymbol{g}^{(Z)}(x, y)$ defined by Eq. (8).

The plane-waves can be expressed by superposition of the cylindrical-waves associated with the Bessel functions, and the translation relation is

$$\boldsymbol{f}_2^{(\pm)}(x, y; \xi) = \boldsymbol{A}^{(\pm)}(\xi)\, \boldsymbol{g}^{(J)}(x, y) \tag{13}$$

$$\left(\boldsymbol{A}^{(\pm)}(\xi)\right)_{n,m} = \left(\frac{i\,\alpha_n(\xi) \pm \beta_{2,n}(\xi)}{k_s}\right)^m . \tag{14}$$

Also, we have the following relation:

$$\sum_{l\in\mathbb{Z}} \boldsymbol{g}^{(H^{(1)})}(x - l\,d, y)\, e^{i\,l\,d\,\xi} = \begin{cases} \boldsymbol{B}^{(+)}(\xi)\, \boldsymbol{f}_2^{(+)}(x, y; \xi) & \text{for } y > 0 \\ \boldsymbol{B}^{(-)}(\xi)\, \boldsymbol{f}_2^{(-)}(x, y; \xi) & \text{for } y < 0 \end{cases} \tag{15}$$

$$\left(\boldsymbol{B}^{(\pm)}(\xi)\right)_{n,m} = \frac{2}{d\,\beta_{2,m}(\xi)} \left(\frac{-i\,\alpha_m(\xi) \pm \beta_{2,m}(\xi)}{k_2}\right)^n . \tag{16}$$

4 Outline of Formulation

4.1 Integral Equation for Cylinder Array

The PPFT fields outside the cylinders (dashed circles) can be expressed in the plane-wave expansions as

$$\overline{\psi}(x; \xi, y) = \boldsymbol{f}_r^{(-)}(x, y - y'; \xi)^t\, \overline{\boldsymbol{\psi}}_r^{(-)}(\xi, y') + \boldsymbol{f}_r^{(+)}(x, y - y'; \xi)^t\, \overline{\boldsymbol{\psi}}_r^{(+)}(\xi, y') \tag{17}$$

for $r = 1, 2, 3$, where $\overline{\boldsymbol{\psi}}_r^{(\pm)}(\xi, y')$ denotes the column matrix containing the amplitude of the plane-waves propagating in the $\pm y$-direction at $y = y'$. The incident field for the cylinder array consists of the waves propagating in the $-y$-direction from the plane $y = h_1$ and the waves propagating in the $+y$-direction from the plane $y = h_2$. Therefore, the incident field transformed by the PPFT is given by

$$\overline{\psi}^{(i)}(x; \xi, y) = \boldsymbol{f}_2^{(-)}(x, y - h_1; \xi)^t\, \overline{\boldsymbol{\psi}}_2^{(-)}(\xi, h_1 - 0) + \boldsymbol{f}_2^{(+)}(x, y - h_2; \xi)^t\, \overline{\boldsymbol{\psi}}_2^{(+)}(\xi, h_2 + 0). \tag{18}$$

Using Eqs. (6), (10), and (13), the PPFT incident field is also expressed in the cylindrical-wave expansion as

$$\overline{\psi}^{(i)}(x; \xi, y) = \boldsymbol{g}^{(J)}(x - q\,d, y)^t\, \overline{\boldsymbol{a}}^{(i)}(\xi)\, e^{i\,q\,d\,\xi} \tag{19}$$

where the expansion coefficient matrix $\overline{\boldsymbol{a}}^{(i)}(\xi)$ is derived as

$$\overline{\boldsymbol{a}}^{(i)}(\xi) = \boldsymbol{A}^{(-)}(\xi)^t\, \boldsymbol{F}_2(h_1; \xi)\, \overline{\boldsymbol{\psi}}_2^{(-)}(\xi, h_1 - 0) + \boldsymbol{A}^{(+)}(\xi)^t\, \boldsymbol{F}_2(-h_2; \xi)\, \overline{\boldsymbol{\psi}}_2^{(+)}(\xi, h_2 + 0). \tag{20}$$

The inverse transform given in Eq. (2) is applied, and we may obtain the following expression:

$$\psi^{(i)}(x, y) = \boldsymbol{g}^{(J)}(x - q\,d, y)^t\, \boldsymbol{a}_q^{(i)} \tag{21}$$

$$\boldsymbol{a}_q^{(i)} = \frac{1}{k_d} \int_{-k_d/2}^{k_d/2} \overline{\boldsymbol{a}}^{(i)}(\xi)\, e^{i\,q\,d\,\xi}\, \mathrm{d}\xi. \tag{22}$$

On the other hand, the scattered field from the cylinder array consists of the outward propagating waves from each cylinder, and it can be expressed as

$$\psi^{(s)}(x, y) = \sum_{q \in \mathbb{Z}} \boldsymbol{g}^{(H^{(1)})}(x - q\,d, y)^t\, \boldsymbol{a}_q^{(s)} \tag{23}$$

where $\boldsymbol{a}_q^{(s)}$ denotes the column matrix generated by the expansion coefficients of the outward propagating cylindrical-waves from the qth-cylinder and $\boldsymbol{a}_q^{(s)} = \boldsymbol{0}$ for $q \in \mathcal{D}$. Using Eq. (12), the total field near but outside the qth-cylinder (dashed circle in Fig. 1 for $q \in \mathcal{D}$) is expressed as

$$\begin{aligned} \psi(x, y) = \boldsymbol{g}^{(J)}(x - q\,d, y)^t & \left(\boldsymbol{a}_q^{(i)} + \sum_{q' \in \mathbb{Z} \setminus \{q\}} \boldsymbol{G}^{(H^{(1)})}(\left(q - q'\right) d, 0)^t\, \boldsymbol{a}_{q'}^{(s)} \right) \\ & + \boldsymbol{g}^{(H^{(1)})}(x - q\,d, y)^t\, \boldsymbol{a}_q^{(s)} \end{aligned} \tag{24}$$

where $\mathbb{Z} \setminus \{q\}$ denotes the set $\{q\}$ is removed from the integer set $\mathbb{Z}$. The first and the second terms on the right-hand side of Eq. (24) are given by superpositions of cylindrical-waves associated with the Bessel function and the Hankel function of the first kind, respectively, and they represent the incident and the scattered fields for the qth-cylinder. Since the relation between the coefficient matrices of the incident and the scattered fields are given by the transition-matrix, we have

$$\boldsymbol{a}_q^{(s)} = \begin{cases} \boldsymbol{T} \left(\boldsymbol{a}_q^{(i)} + \sum_{q' \in \mathbb{Z} \setminus \{q\}} \boldsymbol{G}^{(H^{(1)})}(\left(q - q'\right) d, 0)^t\, \boldsymbol{a}_{q'}^{(s)} \right) & \text{for } q \in \mathbb{Z} \setminus \mathcal{D} \\ \boldsymbol{0} & \text{for } q \in \mathcal{D} \end{cases} \tag{25}$$

where the transition-matrix $\boldsymbol{T}$ is a diagonal matrix whose (n, m)-components are given by

$$(\boldsymbol{T})_{n,m} = \begin{cases} \delta_{n,m}\, \frac{\zeta_2\, J_n(k_2\, a_p)\, J_n'(k_p\, a_p) - \zeta_p\, J_n'(k_2\, a_p)\, J_n(k_p\, a_p)}{\zeta_p\, H_n^{(1)\prime}(k_2\, a_p)\, J_n(k_p\, a_p) - \zeta_2\, H_n^{(1)}(k_2\, a_p)\, J_n'(k_p\, a_p)} & \text{for TM-pol.} \\ \delta_{n,m}\, \frac{\zeta_p\, J_n(k_2\, a_p)\, J_n'(k_p\, a_p) - \zeta_2\, J_n'(k_2\, a_p)\, J_n(k_p\, a_p)}{\zeta_2\, H_n^{(1)\prime}(k_2\, a_p)\, J_n(k_p\, a_p) - \zeta_p\, H_n^{(1)}(k_2\, a_p)\, J_n'(k_p\, a_p)} & \text{for TE-pol.} \end{cases}. \tag{26}$$

The PPFT defined by Eq. (1) is applied to Eq. (23), and the PPFT scattered field is expressed as

$$\overline{\psi}^{(s)}(x; \xi, y) = \sum_{m \in \mathbb{Z}} \boldsymbol{g}^{(H^{(1)})}(x - m\,d, y)^t\, \overline{\boldsymbol{a}}^{(s)}(\xi)\, e^{i\,m\,d\,\xi} \tag{27}$$

with

$$\overline{\boldsymbol{a}}^{(s)}(\xi) = \sum_{q \in \mathbb{Z}} \boldsymbol{a}_q^{(s)} \, e^{-i\,q\,d\,\xi}. \tag{28}$$

Substituting Eq. (25) in to Eq. (28), we may derive the following integral equation:

$$\begin{aligned} \overline{\boldsymbol{a}}^{(s)}(\xi) &= \boldsymbol{T} \left(\overline{\boldsymbol{a}}^{(i)}(\xi) + \boldsymbol{L}(\xi)\, \overline{\boldsymbol{a}}^{(s)}(\xi) \right) \\ &\quad + \frac{1}{k_d} \boldsymbol{T} \int_{-k_d/2}^{k_d/2} c(\xi, \xi') \left(\overline{\boldsymbol{a}}^{(i)}(\xi') + \boldsymbol{L}(\xi')\, \overline{\boldsymbol{a}}^{(s)}(\xi') \right) \mathrm{d}\xi' \end{aligned} \tag{29}$$

with

$$\boldsymbol{L}(\xi) = \sum_{q \in \mathbb{Z} \setminus \{0\}} \boldsymbol{G}^{(H^{(1)})}(q\,d, 0)^t \, e^{-i\,q\,d\,\xi} \tag{30}$$

$$c(\xi, \xi') = -\sum_{q \in \mathcal{D}} e^{-i\,q\,d\,(\xi - \xi')}. \tag{31}$$

The entries of matrix $\boldsymbol{L}(\xi)$ are the lattice sums, and the scalar function $c(\xi, \xi')$ is given by a sum of finite terms, which expresses the effects of defects in the periodic array.

4.2 Field Decomposition

The column matrices of the plane-wave amplitudes other than the ones corresponding to the incoming waves in the upper and the lower regions are here expressed by a sum of two column matrices as

$$\overline{\boldsymbol{\psi}}_r^{(\pm)}(\xi, y') = \overline{\boldsymbol{\psi}}_r^{(\pm,p)}(\xi, y') + \overline{\boldsymbol{\psi}}_r^{(\pm,d)}(\xi, y') \tag{32}$$

for $r = 1, 2, 3$. The first matrices $\overline{\boldsymbol{\psi}}_r^{(\pm,p)}(\xi, y')$ correspond to the scattered field by the perfectly periodic cylinder array without defects, and the second matrices $\overline{\boldsymbol{\psi}}_r^{(\pm,d)}(\xi, y')$ correspond to the residual field. Then the coefficient matrices $\overline{\boldsymbol{a}}^{(f)}(\xi)$ $(f = i, s)$ are also decomposed as

$$\overline{\boldsymbol{a}}^{(f)}(\xi) = \overline{\boldsymbol{a}}^{(f,p)}(\xi) + \overline{\boldsymbol{a}}^{(f,d)}(\xi) \tag{33}$$

where $\overline{\boldsymbol{a}}^{(f,p)}(\xi)$ and $\overline{\boldsymbol{a}}^{(f,d)}(\xi)$ correspond to the scattered field by the perfectly periodic cylinder array $\psi^{(s,p)}(x, y)$ and the residual field $\psi^{(s,d)}(x, y)$, respectively.

From Eqs. (15) and (23), we may obtain the plane-wave expansion representations of the decomposed scattered fields inside the slab region as

$$\overline{\psi}^{(s,u)}(x;\xi,y) = \begin{cases} \boldsymbol{f}_2^{(+)}(x,y;\xi)^t\, \overline{\boldsymbol{\psi}}_2^{(+,u)}(\xi,+0) & \text{for } 0 < y < h_1 \\ \boldsymbol{f}_2^{(-)}(x,y;\xi)^t\, \overline{\boldsymbol{\psi}}_2^{(-,u)}(\xi,-0) & \text{for } h_2 < y < 0 \end{cases} \tag{34}$$

for $u = p, d$, where the decomposed amplitudes are derived as

$$\overline{\boldsymbol{\psi}}_2^{(\pm,u)}(\xi,\pm 0) = \boldsymbol{B}^{(\pm)}(\xi)^t\, \overline{\boldsymbol{a}}^{(s,u)}(\xi). \tag{35}$$

Equation (20) yields

$$\begin{aligned} \overline{\boldsymbol{a}}^{(i,u)}(\xi) = {} & \boldsymbol{A}^{(-)}(\xi)^t\, \boldsymbol{F}_2(h_1;\xi)\, \overline{\boldsymbol{\psi}}_2^{(-,u)}(\xi,h_1-0) \\ & + \boldsymbol{A}^{(+)}(\xi)^t\, \boldsymbol{F}_2(-h_2;\xi)\, \overline{\boldsymbol{\psi}}_2^{(+,u)}(\xi,h_2+0), \end{aligned} \tag{36}$$

and, from the boundary conditions at $y = h_r$ $(r = 1, 2)$ and Eq. (10), we obtain the following relations:

$$\begin{pmatrix} \overline{\boldsymbol{\psi}}_1^{(+,p)}(\xi,h_1+0) \\ \overline{\boldsymbol{\psi}}_2^{(-,p)}(\xi,h_1-0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{11}^{(1)}(\xi) & \boldsymbol{S}_{12}^{(1)}(\xi)\, \boldsymbol{F}_2(h_1;\xi) \\ \boldsymbol{S}_{21}^{(1)}(\xi) & \boldsymbol{S}_{22}^{(1)}(\xi)\, \boldsymbol{F}_2(h_1;\xi) \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{\psi}}_1^{(-)}(\xi,h_1+0) \\ \overline{\boldsymbol{\psi}}_2^{(+,p)}(\xi,+0) \end{pmatrix} \tag{37}$$

$$\begin{pmatrix} \overline{\boldsymbol{\psi}}_2^{(+,p)}(\xi,h_2+0) \\ \overline{\boldsymbol{\psi}}_3^{(-,p)}(\xi,h_2-0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{11}^{(2)}(\xi)\, \boldsymbol{F}_2(-h_2;\xi) & \boldsymbol{S}_{12}^{(2)}(\xi) \\ \boldsymbol{S}_{21}^{(2)}(\xi)\, \boldsymbol{F}_2(-h_2;\xi) & \boldsymbol{S}_{22}^{(2)}(\xi) \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{\psi}}_2^{(-,p)}(\xi,-0) \\ \overline{\boldsymbol{\psi}}_3^{(+)}(\xi,h_2-0) \end{pmatrix} \tag{38}$$

$$\begin{pmatrix} \overline{\boldsymbol{\psi}}_1^{(+,d)}(\xi,h_1+0) \\ \overline{\boldsymbol{\psi}}_2^{(-,d)}(\xi,h_1-0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{12}^{(1)}(\xi)\, \boldsymbol{F}_2(h_1;\xi) \\ \boldsymbol{S}_{22}^{(1)}(\xi)\, \boldsymbol{F}_2(h_1;\xi) \end{pmatrix} \overline{\boldsymbol{\psi}}_2^{(+,d)}(\xi,+0) \tag{39}$$

$$\begin{pmatrix} \overline{\boldsymbol{\psi}}_2^{(+,d)}(\xi,h_2+0) \\ \overline{\boldsymbol{\psi}}_3^{(-,d)}(\xi,h_2-0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{11}^{(2)}(\xi)\, \boldsymbol{F}_2(-h_2;\xi) \\ \boldsymbol{S}_{21}^{(2)}(\xi)\, \boldsymbol{F}_2(-h_2;\xi) \end{pmatrix} \overline{\boldsymbol{\psi}}_2^{(-,d)}(\xi,-0) \tag{40}$$

$$\left(\boldsymbol{S}_{11}^{(r)}(\xi)\right)_{n,m} = \begin{cases} -\delta_{n,m}\, \frac{\mu_r\, \beta_{r+1,m}(\xi) - \mu_{r+1}\, \beta_{r,m}(\xi)}{\mu_r\, \beta_{r+1,n}(\xi) + \mu_{r+1}\, \beta_{r,n}(\xi)} & \text{for TM-pol.} \\ -\delta_{n,m}\, \frac{\varepsilon_r\, \beta_{r+1,m}(\xi) - \varepsilon_{r+1}\, \beta_{r,m}(\xi)}{\varepsilon_r\, \beta_{r+1,n}(\xi) + \varepsilon_{r+1}\, \beta_{r,n}(\xi)} & \text{for TE-pol.} \end{cases} \tag{41}$$

$$\boldsymbol{S}_{12}^{(r)}(\xi) = \boldsymbol{I} - \boldsymbol{S}_{11}^{(r)}(\xi), \qquad \boldsymbol{S}_{21}^{(r)}(\xi) = \boldsymbol{I} + \boldsymbol{S}_{11}^{(r)}(\xi), \qquad \boldsymbol{S}_{22}^{(r)}(\xi) = -\boldsymbol{S}_{11}^{(r)}(\xi) \tag{42}$$

where $\boldsymbol{I}$ denotes the identity matrix.

4.3 Scattering by Periodic Cylinder Array

When we consider the scattering from the perfectly periodic cylinder array without defects, the second terms on the right-hand side of Eq. (29) vanishes, and we obtain

$$\overline{\boldsymbol{a}}^{(s,p)}(\xi) = (\boldsymbol{I} - \boldsymbol{T}\, \boldsymbol{L}(\xi))^{-1}\, \boldsymbol{T}\, \overline{\boldsymbol{a}}^{(i,p)}(\xi). \tag{43}$$

Then, using Eqs. (35) and (36), the input-output relation for the periodic cylinder array is derived as

$$\begin{pmatrix} \overline{\boldsymbol{\psi}}_2^{(+,p)}(\xi,+0) \\ \overline{\boldsymbol{\psi}}_2^{(-,p)}(\xi,-0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{11}^{(c)}(\xi) & \boldsymbol{S}_{12}^{(c)}(\xi) \\ \boldsymbol{S}_{21}^{(c)}(\xi) & \boldsymbol{S}_{22}^{(c)}(\xi) \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{\psi}}_2^{(-,p)}(\xi,h_1-0) \\ \overline{\boldsymbol{\psi}}_2^{(+,p)}(\xi,h_2+0) \end{pmatrix} \tag{44}$$

$$\boldsymbol{S}_{11}^{(c)}(\xi) = \boldsymbol{B}^{(+)}(\xi)^t\,(\boldsymbol{I}-\boldsymbol{T}\,\boldsymbol{L}(\xi))^{-1}\,\boldsymbol{T}\,\boldsymbol{A}^{(-)}(\xi)^t\,\boldsymbol{F}_2(h_1;\xi) \tag{45}$$

$$\boldsymbol{S}_{12}^{(c)}(\xi) = \boldsymbol{B}^{(+)}(\xi)^t\,(\boldsymbol{I}-\boldsymbol{T}\,\boldsymbol{L}(\xi))^{-1}\,\boldsymbol{T}\,\boldsymbol{A}^{(+)}(\xi)^t\,\boldsymbol{F}_2(-h_2;\xi) \tag{46}$$

$$\boldsymbol{S}_{21}^{(c)}(\xi) = \boldsymbol{B}^{(-)}(\xi)^t\,(\boldsymbol{I}-\boldsymbol{T}\,\boldsymbol{L}(\xi))^{-1}\,\boldsymbol{T}\,\boldsymbol{A}^{(-)}(\xi)^t\,\boldsymbol{F}_2(h_1;\xi) \tag{47}$$

$$\boldsymbol{S}_{22}^{(c)}(\xi) = \boldsymbol{B}^{(-)}(\xi)^t\,(\boldsymbol{I}-\boldsymbol{T}\,\boldsymbol{L}(\xi))^{-1}\,\boldsymbol{T}\,\boldsymbol{A}^{(+)}(\xi)^t\,\boldsymbol{F}_2(-h_2;\xi). \tag{48}$$

From Eqs. (37), (38), and (44), the plane-wave amplitudes for the scattered field by the perfectly periodic cylinder array without defects are obtained as follows:

$$\begin{pmatrix} \overline{\boldsymbol{\psi}}_1^{(+,p)}(\xi,h_1+0) \\ \overline{\boldsymbol{\psi}}_3^{(-,p)}(\xi,h_2-0) \\ \overline{\boldsymbol{\psi}}_2^{(-,p)}(\xi,h_1-0) \\ \overline{\boldsymbol{\psi}}_2^{(+,p)}(\xi,h_2+0) \\ \overline{\boldsymbol{\psi}}_2^{(+,p)}(\xi,+0) \\ \overline{\boldsymbol{\psi}}_2^{(-,p)}(\xi,-0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{11}^{(p)}(\xi) & \boldsymbol{S}_{12}^{(p)}(\xi) \\ \boldsymbol{S}_{21}^{(p)}(\xi) & \boldsymbol{S}_{22}^{(p)}(\xi) \\ \boldsymbol{S}_{31}^{(p)}(\xi) & \boldsymbol{S}_{32}^{(p)}(\xi) \\ \boldsymbol{S}_{41}^{(p)}(\xi) & \boldsymbol{S}_{42}^{(p)}(\xi) \\ \boldsymbol{S}_{51}^{(p)}(\xi) & \boldsymbol{S}_{52}^{(p)}(\xi) \\ \boldsymbol{S}_{61}^{(p)}(\xi) & \boldsymbol{S}_{62}^{(p)}(\xi) \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{\psi}}_1^{(-)}(\xi,h_1+0) \\ \overline{\boldsymbol{\psi}}_3^{(+)}(\xi,h_2-0) \end{pmatrix} \tag{49}$$

$$\boldsymbol{M}_1^{(p)}(\xi) = \Big(\boldsymbol{I} - \boldsymbol{S}_{11}^{(2)}(\xi)\,\boldsymbol{F}_2(-h_2;\xi)\,\boldsymbol{S}_{22}^{(c)}(\xi)\Big)^{-1} \tag{50}$$

$$\boldsymbol{M}_2^{(p)}(\xi) = \boldsymbol{S}_{11}^{(c)}(\xi) + \boldsymbol{S}_{12}^{(c)}(\xi)\,\boldsymbol{M}_1^{(p)}(\xi)\,\boldsymbol{S}_{11}^{(2)}(\xi)\,\boldsymbol{F}_2(-h_2;\xi)\,\boldsymbol{S}_{21}^{(c)}(\xi) \tag{51}$$

$$\boldsymbol{M}_3^{(p)}(\xi) = \Big(\boldsymbol{I} - \boldsymbol{S}_{22}^{(1)}(\xi)\,\boldsymbol{F}_2(h_1;\xi)\,\boldsymbol{M}_2^{(p)}(\xi)\Big)^{-1} \tag{52}$$

$$\boldsymbol{S}_{31}^{(p)}(\xi) = \boldsymbol{M}_3^{(p)}(\xi)\,\boldsymbol{S}_{21}^{(1)}(\xi) \tag{53}$$

$$\boldsymbol{S}_{32}^{(p)}(\xi) = \boldsymbol{M}_3^{(p)}(\xi)\,\boldsymbol{S}_{22}^{(1)}(\xi)\,\boldsymbol{F}_2(h_1;\xi)\,\boldsymbol{S}_{12}^{(c)}(\xi)\,\boldsymbol{M}_1^{(p)}(\xi)\,\boldsymbol{S}_{12}^{(2)}(\xi) \tag{54}$$

$$\boldsymbol{S}_{41}^{(p)}(\xi) = \boldsymbol{M}_1^{(p)}(\xi)\,\boldsymbol{S}_{11}^{(2)}(\xi)\,\boldsymbol{F}_2(-h_2;\xi)\,\boldsymbol{S}_{21}^{(c)}(\xi)\,\boldsymbol{S}_{31}^{(p)}(\xi) \tag{55}$$

$$\boldsymbol{S}_{42}^{(p)}(\xi) = \boldsymbol{M}_1^{(p)}(\xi)\,\Big(\boldsymbol{S}_{11}^{(2)}(\xi)\,\boldsymbol{F}_2(-h_2;\xi)\,\boldsymbol{S}_{21}^{(c)}(\xi)\,\boldsymbol{S}_{32}^{(p)}(\xi) + \boldsymbol{S}_{12}^{(2)}(\xi)\Big) \tag{56}$$

$$\boldsymbol{S}_{51}^{(p)}(\xi) = \boldsymbol{S}_{11}^{(c)}(\xi)\,\boldsymbol{S}_{31}^{(p)}(\xi) + \boldsymbol{S}_{12}^{(c)}(\xi)\,\boldsymbol{S}_{41}^{(p)}(\xi) \tag{57}$$

$$\boldsymbol{S}_{52}^{(p)}(\xi) = \boldsymbol{S}_{11}^{(c)}(\xi)\,\boldsymbol{S}_{32}^{(p)}(\xi) + \boldsymbol{S}_{12}^{(c)}(\xi)\,\boldsymbol{S}_{42}^{(p)}(\xi) \tag{58}$$

$$\boldsymbol{S}_{61}^{(p)}(\xi) = \boldsymbol{S}_{21}^{(c)}(\xi)\,\boldsymbol{S}_{31}^{(p)}(\xi) + \boldsymbol{S}_{22}^{(c)}(\xi)\,\boldsymbol{S}_{41}^{(p)}(\xi) \tag{59}$$

$$\boldsymbol{S}_{62}^{(p)}(\xi) = \boldsymbol{S}_{21}^{(c)}(\xi)\,\boldsymbol{S}_{32}^{(p)}(\xi) + \boldsymbol{S}_{22}^{(c)}(\xi)\,\boldsymbol{S}_{42}^{(p)}(\xi) \tag{60}$$

$$\boldsymbol{S}_{11}^{(p)}(\xi) = \boldsymbol{S}_{11}^{(1)}(\xi) + \boldsymbol{S}_{12}^{(1)}(\xi)\,\boldsymbol{F}_2(h_1;\xi)\,\boldsymbol{S}_{51}^{(p)}(\xi) \tag{61}$$

$$\boldsymbol{S}_{12}^{(p)}(\xi) = \boldsymbol{S}_{12}^{(1)}(\xi)\,\boldsymbol{F}_2(h_1;\xi)\,\boldsymbol{S}_{52}^{(p)}(\xi) \tag{62}$$

$$\boldsymbol{S}_{21}^{(p)}(\xi) = \boldsymbol{S}_{21}^{(2)}(\xi)\,\boldsymbol{F}_2(-h_2;\xi)\,\boldsymbol{S}_{61}^{(p)}(\xi) \tag{63}$$

$$\boldsymbol{S}_{22}^{(p)}(\xi) = \boldsymbol{S}_{21}^{(2)}(\xi)\,\boldsymbol{F}_2(-h_2;\xi)\,\boldsymbol{S}_{62}^{(p)}(\xi) + \boldsymbol{S}_{22}^{(2)}(\xi). \tag{64}$$

Also, Eq. (36) yields the following expression of the expansion coefficient for the incident field:

$$\overline{\boldsymbol{a}}^{(i,p)}(\xi) = \boldsymbol{Q}_1^{(p)}\, \overline{\boldsymbol{\psi}}_1^{(-)}(\xi, h_1 + 0) + \boldsymbol{Q}_2^{(p)}\, \overline{\boldsymbol{\psi}}_3^{(+)}(\xi, h_2 - 0) \tag{65}$$

$$\boldsymbol{Q}_r^{(p)} = \boldsymbol{A}^{(-)}(\xi)^t\, \boldsymbol{F}_2(h_1; \xi)\, \boldsymbol{S}_{3r}^{(p)}(\xi) + \boldsymbol{A}^{(+)}(\xi)^t\, \boldsymbol{F}_2(-h_2; \xi)\, \boldsymbol{S}_{4r}^{(p)}(\xi) \tag{66}$$

for $r = 1, 2$.

4.4 Residual Field

Using Eqs. (36), (35), (39), and (40), we obtain the following relation:

$$\overline{\boldsymbol{a}}^{(i,d)}(\xi) = \boldsymbol{P}(\xi)\, \overline{\boldsymbol{a}}^{(s,d)}(\xi) \tag{67}$$

$$\begin{aligned}\boldsymbol{P}(\xi) = {} & \boldsymbol{A}^{(-)}(\xi)^t\, \boldsymbol{F}_2(h_1; \xi)\, \boldsymbol{S}_{22}^{(1)}(\xi)\, \boldsymbol{F}_2(h_1; \xi)\, \boldsymbol{B}^{(+)}(\xi)^t \\ & + \boldsymbol{A}^{(+)}(\xi)^t\, \boldsymbol{F}_2(-h_2; \xi)\, \boldsymbol{S}_{11}^{(2)}(\xi)\, \boldsymbol{F}_2(-h_2; \xi)\, \boldsymbol{B}^{(-)}(\xi)^t.\end{aligned} \tag{68}$$

Then, from Eqs. (29), (33), (43), and (67), we may understand the expansion coefficient for the residual field satisfies the following integral-equation:

$$\begin{aligned}& [\boldsymbol{I} - \boldsymbol{T}\, (\boldsymbol{L}(\xi) + \boldsymbol{P}(\xi))]\, \overline{\boldsymbol{a}}^{(s,d)}(\xi) \\ & \quad - \frac{1}{k_d}\, \boldsymbol{T} \int_{-k_d/2}^{k_d/2} c(\xi, \xi')\, \big(\boldsymbol{L}(\xi') + \boldsymbol{P}(\xi')\big)\, \overline{\boldsymbol{a}}^{(s,d)}(\xi')\, \mathrm{d}\xi' = \boldsymbol{q}(\xi)\end{aligned} \tag{69}$$

where the right-hand side is given by

$$\boldsymbol{q}(\xi) = \frac{1}{k_d}\, \boldsymbol{T} \int_{-k_d/2}^{k_d/2} c(\xi, \xi')\, \big(\boldsymbol{I} - \boldsymbol{L}(\xi')\, \boldsymbol{T}\big)^{-1}\, \overline{\boldsymbol{a}}^{(i,p)}(\xi')\, \mathrm{d}\xi'. \tag{70}$$

To solve the integral equation (69), we introduce a discretization in the transform parameter ξ. Since the column matrix $\overline{\boldsymbol{a}}^{(s,d)}(\xi)$ is a periodic function with the period k_d, we take L sample points $\{\xi_l\}_{l=1}^{L}$ in the Brillouin zone $-k_d/2 < \xi \leq k_d$ and assume that Eq. (69) is satisfied at these points. Also the integral on the left-hand side is approximated by an appropriate numerical integration scheme with the use of same sample points. We denote the respective weights of the sample points by $\{w_l\}_{l=1}^{L}$. Then, we may derive the following relation:

$$\sum_{l'=1}^{L} \boldsymbol{C}_{l,l'}\, \overline{\boldsymbol{a}}^{(s,d)}(\xi_{l'}) = \boldsymbol{q}(\xi_l) \tag{71}$$

with

$$\boldsymbol{C}_{l,l'} = -\left(\frac{w_{l'}}{k_d}\,c(\xi_l,\xi_{l'}) + \delta_{l,l'}\right)\boldsymbol{T}\,(\boldsymbol{L}(\xi_{l'}) + \boldsymbol{P}(\xi_{l'})) + \delta_{l,l'}\,\boldsymbol{I}. \tag{72}$$

From Eq. (71), the coefficient matrices of the residual field at the sample point is obtained as

$$\begin{pmatrix} \overline{\boldsymbol{a}}^{(s,d)}(\xi_1) \\ \vdots \\ \overline{\boldsymbol{a}}^{(s,d)}(\xi_L) \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}_{1,1} & \cdots & \boldsymbol{C}_{1,L} \\ \vdots & \ddots & \vdots \\ \boldsymbol{C}_{L,1} & \cdots & \boldsymbol{C}_{L,L} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{q}(\xi_1) \\ \vdots \\ \boldsymbol{q}(\xi_L) \end{pmatrix}. \tag{73}$$

From Eqs. (35), (39), and (40), the plane-wave amplitudes for the residual field at the sample points are obtained by

$$\begin{pmatrix} \overline{\boldsymbol{\psi}}_1^{(+,d)}(\xi_l, h_1+0) \\ \overline{\boldsymbol{\psi}}_3^{(-,d)}(\xi_l, h_2-0) \\ \overline{\boldsymbol{\psi}}_2^{(-,d)}(\xi_l, h_1-0) \\ \overline{\boldsymbol{\psi}}_2^{(+,d)}(\xi_l, h_2+0) \\ \overline{\boldsymbol{\psi}}_2^{(+,d)}(\xi_l, +0) \\ \overline{\boldsymbol{\psi}}_2^{(-,d)}(\xi_l, -0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{12}^{(1)}(\xi_l)\,\boldsymbol{F}_2(h_1;\xi_l)\,\boldsymbol{B}^{(+)}(\xi_l)^t \\ \boldsymbol{S}_{21}^{(2)}(\xi_l)\,\boldsymbol{F}_2(-h_2;\xi_l)\,\boldsymbol{B}^{(-)}(\xi_l)^t \\ \boldsymbol{S}_{22}^{(1)}(\xi_l)\,\boldsymbol{F}_2(h_1;\xi_l)\,\boldsymbol{B}^{(+)}(\xi_l)^t \\ \boldsymbol{S}_{11}^{(2)}(\xi_l)\,\boldsymbol{F}_2(-h_2;\xi_l)\,\boldsymbol{B}^{(-)}(\xi_l)^t \\ \boldsymbol{B}^{(+)}(\xi_l)^t \\ \boldsymbol{B}^{(-)}(\xi_l)^t \end{pmatrix} \overline{\boldsymbol{a}}^{(s,d)}(\xi_l). \tag{74}$$

5 Concluding Remarks

This paper presents a formulation of the electromagnetic scattering from the defected periodic cylinder array embedded in dielectric slab. The present formulation is based on the RTMA with the help of the PPFT, and the discretization for the numerical computation is introduced on the transform parameter ξ, which is related to the wavenumber. The main problem on the spectral-domain analysis is generally summarized to the discretization scheme on the wavenumber space. In the present formulation, the discretization scheme can be considered only inside the Brillouin zone owing to the PPFT.

References

1. Watanabe, K., Yasumoto, K.: Two-Dimensional Electromagnetic Scattering of Non-Plane Incident Waves by Periodic Structures. Progress in Electromagnetics Res., PIER **74**, 241–271 (2007)
2. Petit, R. (ed.): Electromagnetic Theory of Gratings. Springer, Berlin (1980)
3. Chew, W.C.: Waves and Fields in Inhomogeneous Media. Van Nostrand Reinhold, New York (1990)
4. Felbacq, D., Tayeb, G., Maystre, D.: Scattering by a Random Set of Parallel Cylinders. J. Opt. Soc. Am. A **11**, 2526–2538 (1994)
5. Roussel, H., Chew, W.C., Jouvie, F., Tabbara, W.: Electromagnetic Scattering from Dielectric and Magnetic Gratings of Fibers – a T-Matrix Solution. J. Electromagn. Waves Appl. **10**, 109–127 (1996)

Numerical Examination on Transmission Properties of FBG by FDTD Method

Angger Abdul Razak and Mitsuhiro Yokota

Abstract It is important to examine fiber Bragg grating (FBG) for the design of the practical use. Finite difference time domain (FDTD) method is used to examine the transmission properties. FBG is assumed to be two-dimensional periodic structures. Three different grating refractive index shape shows similar reflected frequency, and the square type show the lowest power transmission. Higher grating refractive index value affect in smaller value of transmission amplitude in specific frequency, and the center of transmission frequency drop is slightly shifted to lower value. Correlation between grating period and center of transmission frequency drop is in a nonlinear correlation. Longer total grating length will affect to the lower transmission value, and correlation between them is nearly linear in a certain area. Chirped FBG (CFBG) show a wider broadband reflect compared to the uniform FBG type, with a weaker reflectance. Increasing and decreasing grating period of CFBG show similar transmission characteristic.

Keywords FBG · CFBG · FDTD · Frequency filter

1 Introduction

Optical fibers have made the big revolution for landlines telecommunication systems. One component that takes a part in this technology is fiber Bragg grating (FBG). This device has an important role for the optical amplifier and filter of the optical system. An optical amplifier would not work well without the existence of FBG [1]. FBG can reflect, transmit or filter a certain wavelength based on the design [2]. In optical systems, FBG could be managed as band reject filter and can be used as a dispersion compensator [3-4]. FBG also could be used as a sensor, such as gas sensor [5] or refractive index measurement device [6]. Three-dimensional approach simulation of FBG already simulated [7].

A.A. Razak · M. Yokota(✉)
University of Miyazaki, Gakuenkibanadai Nishi 1-1, Miyazaki 889-2152, Japan
e-mail: angger.abdulrazak@gmail.com, t0b210u@cc.miyazaki-u.ac.jp

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_40

In this paper, FBG is assumed to be modelled by the two-dimensional periodic structures where a single mode can propagate, in order to minimize the computer resource required for simulation process. This letter intends to simulate FBG structure and analyze its output characteristic. Finite difference time domain (FDTD) method is used to simulate and analyze FBG structure [8-10].

2 Basic Properties of FBG

FBG is an optical fiber that has the grating on its core. FBG will reflect some area of wavelength by those grating structures. λ_B is the wavelength reflected by FBG that satisfies Bragg condition shown below (1) [3].

$$\lambda_B = 2\, n_{eff}\, \Lambda \tag{1}$$

Where n_{eff} is an effective refractive index of the fiber core, and Λ is a grating period of FBG.

3 Numerical Simulation and Result

At first stage, uniform FBG is applied and modelled by the two-dimensional periodic waveguide. Physical FBG structure considered here does not differ in thickness of the fiber. FBG structure simulated in this paper is illustrated in Fig. 1.

The transmission properties from the FBG are examined by the FDTD method. For simulation parameter, the analytical region is z=1.75mm times x=35μm, the length of grating structure is up to 1mm, the core width is 3μm. The refractive index of the cladding is n_3=1.46, while core refractive index is n_2=1.48. The cell sizes are set to be $\Delta x=\Delta z$=87.67nm, and the time step size $\Delta t=1.811\text{x}10^{-16}$s is used.

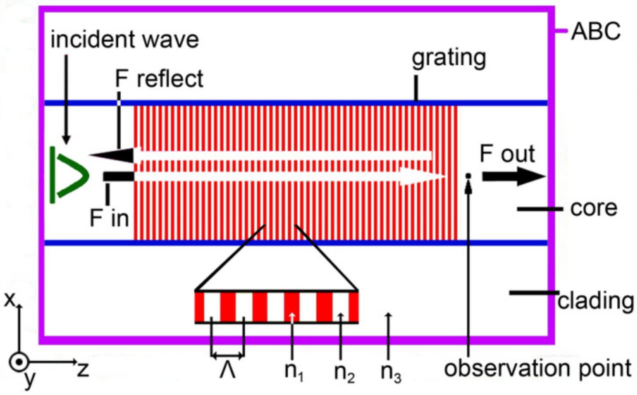
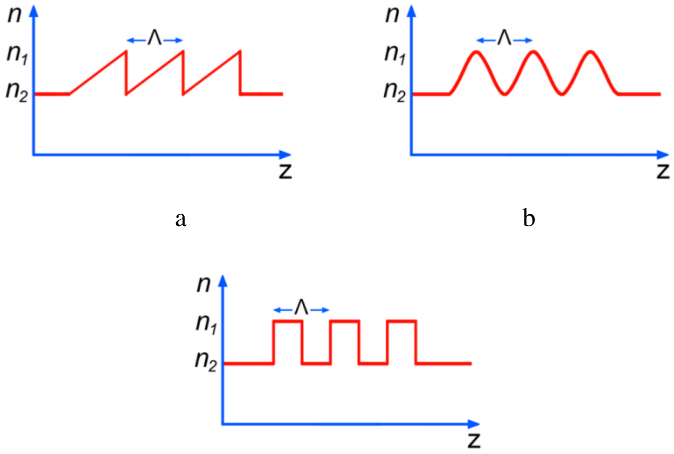


Fig. 1 Illustration of FBG structure simulation area.

The Gaussian pulse is applied as the incident wave in the simulations, which have wavelength center λ_0=1.55μm and corresponding frequency center f_0=1.93x10^{14}Hz. The value of grating period is set as Λ_0=526.19nm. Data collected in time domain will be further processed by using fast Fourier transform (FFT) technique to get output characteristic in frequency domain.

At first, the transmission property of FBG for three different grating refractive index shapes is examined. These types are saw type, sinusoidal type, and square type grating refractive index shapes, respectively as shown in Fig. 2.



a

b

c

Fig. 2 Grating refractive index shape types. a. Saw type grating refractive index shape. b. Sinusoidal type grating refractive index shape. c. Square type grating refractive index shape.

Fig. 3 shows the transmittance in the case of refractive index n_1=1.485. The transmittance is indicated in decibel (dB) as the function of normalized frequency (f/f_0). This results shows that all structure have similar frequency reflected, but they have different transmittance power. Saw type shows the strongest transmittance, while sinusoidal type grating refractive index shape indicate lower power of transmittance in specific frequency, and the lowest transmittance amplitude occur in the square type grating refractive index shape.

The transmission property of FBG for several different values of grating refractive index n_1 is examined in the next simulation to find its effect in output characteristics. Fig. 4 shows the transmittance characteristic results for various grating refractive indexes n_1=1.48, n_1=1.481, n_1=1.482, n_1=1.483, n_1=1.484 and n_1=1.485. When the refractive index n_1 is the same as n_2=1.48, this condition corresponds to no grating structure. The sinusoidal shape type is applied as the grating refractive index type, and grating period Λ=526.19nm.

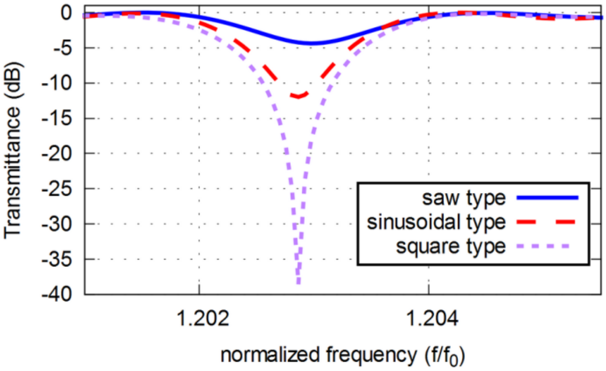


Fig. 3 Transmittance for FBG with different grating refractive index shapes.

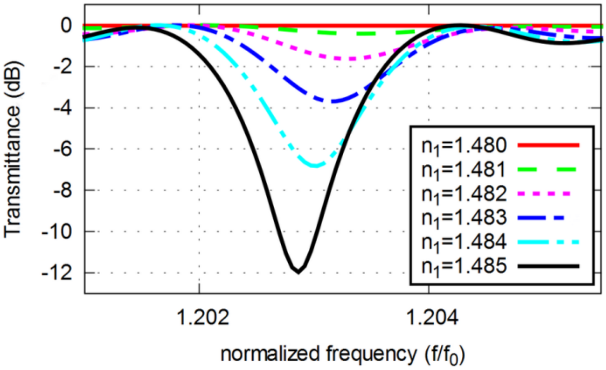


Fig. 4 Transmittance for FBG with different grating refractive index n_1.

By observing Fig. 4, it is shown that if there is no grating (n_1=1.48), FBG transmit all range of frequency in this area. For n_1=1.481, the transmission for some frequency is dropped. While for n_1=1.482 and larger value, we can see different transmission amplitude occur and center of transmission frequency drop also shifted. Comparison between five difference grating refractive indexes shown that higher grating refractive index value affect to smaller value for transmission amplitude, and the center of transmission frequency drops slightly shifted to lower frequency.

After simulating different grating refractive index value, several FBG structures with different grating period are simulated to examine their transmission characteristic properties. For this simulation, refractive index of the cladding is n_3=1.46, core refractive index n_2=1.48, grating refractive index n_1=1.485, and the sinusoidal type grating refractive index shape applied. Grating period structure ranged from $\Lambda=\Lambda_0$ to $\Lambda=1.3\Lambda_0$ with gap of 0.025 Λ_0 for each simulation.

Fig. 5 shows the relation between the center of transmission frequency drop and the normalized grating period. The mark indicates numerical results for every center of transmission frequency drop and the solid line means the nonlinear polynomial regression. Results show that the grating period and the reflected frequency have a negative correlation. Higher grating period will correspond to the lower frequency. From this picture we can observe that correlation between them is not in a linear correlation. This is suits to the Bragg condition.

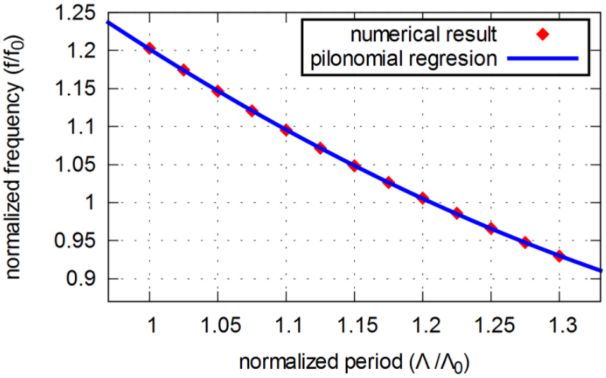


Fig. 5 Center transmission frequency drop of FBG with several grating period Λ.

Afterward, several FBG with different grating length are simulated. Grating length that applied to the FBG is ranged from 0mm (no grating structure) to 1mm, with the differences of 0,1mm for each FBG structures simulated. Grating period Λ=526.19nm and grating refractive index value n_1=1.485 are used.

Fig. 6 show comparison from some result from the simulation. Results show that FBG with gratings structure give similarity in the center of frequency drop. We can see clearly that increasing grating total length will affect to the lower transmission value.

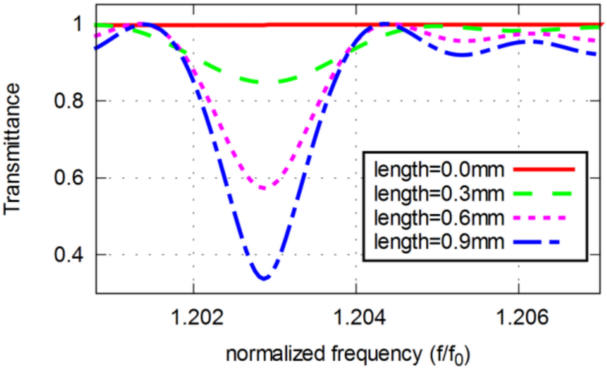


Fig. 6 Transmittance for FBG with different grating length.

Further comparison between all data is shown in Fig. 7. This chart show the correlation between grating length (mm) and transmittance amplitude drop. Each diamond sign indicates numerical results from each simulation. This result show that in the short period, correlation between grating length and transmission drop is not linear, but for grating length longer than 0.2mm, correlation between them is gradually showing a linear condition.

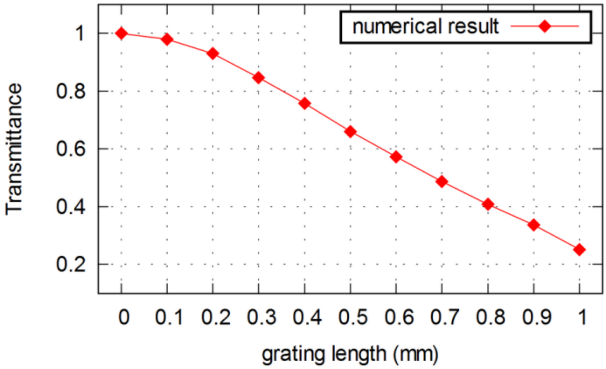


Fig. 7 Transmission frequency drop value of FBG with several grating length.

In the next stage, Chirped Fiber Bragg Grating (CFBG) is simulated. CFBG is a FBG which contain different grating period. Grating period in the CFBG is gradually changing from the first grating toward the end of the grating structures. In this paper, difference period between the first grating structures with the last grating structure is ΔΛ. Illustration of CFBG structures simulated in this research is shown in Fig. 8.

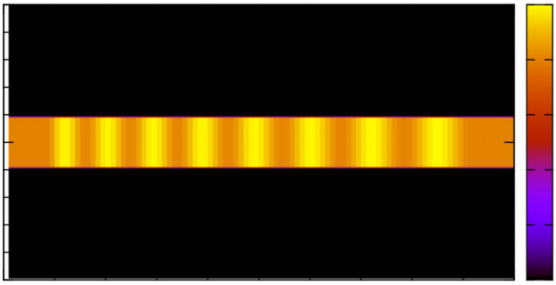


Fig. 8 Illustration of simulation model area for Chirped FBG structure.

For first simulation in CFBG structure, ΔΛ is simulated with several different values. We compare CFBG output characteristic between ΔΛ=0% (uniform FBG), ΔΛ=0.2%, ΔΛ=0.4%, and ΔΛ=0.6%. Result from the simulation is shown in Fig 9. In this pictures, we can see that different ΔΛ in the grating structure will generate different output characteristic. If we increase ΔΛ, wider broadband frequency will be reflected by CFBG, but the reflection power is decreased. This is result is reasonable, because when different period of first and last grating become larger, more variation of grating period will reflect wider number of wavelengths. But if total grating structures length remain the same, increasing ΔΛ will reduce the number of grating period that reflect a specific wavelength, which affect to lower reflectance power.

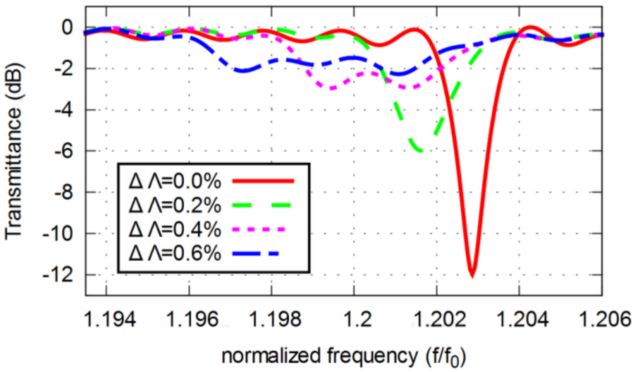


Fig. 9 Transmittance for FBG with different ΔΛ.

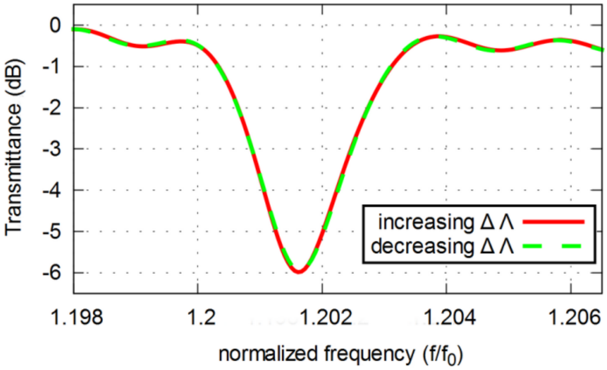


Fig. 10 Transmittance for FBG for increasing and decreasing ΔΛ.

Last simulation for CFBG is comparing increasing changing in grating period structure with decreasing period type. In this simulation, we use similar CFBG structure that have ΔΛ=0.2% but flipped vertically. First CFBG structure have first grating period $\Lambda=\Lambda_0$ and last grating period is $\Lambda=1.002\Lambda_0$. And for second structure, grating period are $\Lambda=1.002\Lambda_0$ and $\Lambda=\Lambda_0$ for first and last grating period respectively.

Fig. 10 show the result from CFBG last simulation. We can see that both structure have a very similar output characteristic. There is a very small differences occurred between two results if we see this result more closely, but overall characteristic is almost identical.

4 Conclusion

Several different FBG structures with different parameters have been simulated by using FDTD Methods. Different grating refractive index shape shows similar reflected frequency, but the transmission of the square shape type indicate the lowest transmission power among three different grating refractive index types. Higher value for grating refractive index affect in smaller value of transmission amplitude in specific frequency, and the center of transmission frequency drop is slightly shifted to lower value. Increasing period of the grating will result in the lower frequency center drop. Correlation between grating period and center of transmission frequency drop is not a linear correlation. Longer length of the grating will result in lower transmission value in the same frequency center. Correlation between them is nonlinear in short length, but it become similarly linear for grating length longer than 0.2mm. Chirped FBG (CFBG) show wider broadband reflect compared to the uniform FBG type, but it show a higher transmittance or weaker reflectance power. Both increasing and decreasing type of CFBG show similarities in center frequency drop, and their transmission characteristic is almost identical.

References

1. Kashyap, R.: Fiber Bragg Grating, 2nd edn. Elsevier, Burlington (2010)
2. Orthonos, A., Kalli, K.: Fiber Bragg grating fundamental and application in telecommunication and sensing. Artech House, Norwood (1999)
3. Riant, I., et al.: Chirped fiber Bragg gratings for WDM chromatic dispersion compensation in multispan 10-Gb/s transmission. IEEE Journal of Selected Topics in Quantum Electronics **5**(5), 1312–1324 (1999)
4. Singh, P.: Dispersion compensation in an optical fiber by using chirp grating. International Journal of Research in Engineering and Technology **3**(7), 506–508 (2014)
5. Khare, A., Singh, J.: Design and study of chirped fiber Bragg grating for sensing of hazardous gases. International Journal of Computer Application **23**(9), 40–43 (2011)
6. Sun, A., Wu, Z.: A hybrid LPG/CFBG for highly sensitive refractive index measurement. Sensor-Open Access Journal, 7318–7325 (2012)
7. Shibayama, J., Oikawa, T., Yamauchi, J., Nakano, H.: Efficient LOD-BOR-FDTD implementation based on a fundamental scheme. IEEE Photonics Technology Letters **24**(11), 957–959 (2012)
8. Kudou, T., Shimizu, K., Takimoto, Y., Ozeki, T.: Bragg grating filter synthesis using Fourier transform with iteration. IEEE Trans. Electron **E83-C**(6), 898–902 (2000)
9. Miyamoto, T., Momoda, M., Yasumoto, K.: Numerical analysis for 3-dimentional optical waveguide with periodic structure using Fourier series expansion method. IEICE Trans. Japan (section J) **J86-C**(6), 591–600 (2003)
10. Tavlove, A.: Computational electrodynamics. Artech House, Norwood (1995)

Numerical Examination on Validity and Usefulness for FDTD Based on Alternative Direction Implicit Method

Tatsuaki Mitsui and Mitsuhiro Yokota

Abstract Recently, the Finite-Difference Time-Domain (FDTD) method is widely used for solving electromagnetic wave problems such as the scattering, diffraction, propagation, coupling etc. As the FDTD method is an explicit type finite difference method, it must be solved under Courant-Friedrich-Lewy (CFL) condition. Therefore, it is assumed that the calculation time increases when an analysis domain is very large or a cell size is very smaller than the wavelength of the electromagnetic wave. In this paper, the validity and usefulness of the Alternative Direction Implicit FDTD (ADI-FDTD) method, which is removed the constraints of CFL condition, are confirmed numerically when ADI-FDTD method is applied to the propagation and the scattering by a cylinder.

Keywords Numerical techniques · FDTD method · ADI method · CFL condition

1 Introduction

Recently, the performance of information processing technique is increasing by the progress of the computers which make easy the analysis of electromagnetic wave for simulation. In order to resolve the issues of the design of the electromagnetic devices, the development of excellent algorithms has been desired.

The Finite-Difference Time-Domain (FDTD) method has developed to solve the Maxwell's equation numerically [1]. Now a day, the FDTD method is very popular method in the electromagnetic society. The FDTD method is based the discretization of the simultaneous partial differential equation in time and space domain.

However, the FDTD method is explicit method type of difference method, so the restricted CFL condition is required to obtain the stable solution. Accordingly, if

---

T. Mitsui · M. Yokota(✉)
Department of Electrical and Electronic Engineering, University of Miyazaki,
1-1 Gakuen Kibanadai Nishi, Miyazaki 889-2192, Japan
e-mail: hc11084@student.miyazaki-u.ac.jp, m.yokota@m.ieice.org

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_41

the analysis object and domain have the microscopic size in order, the long time transient analysis is necessary. To overcome this defect, the implicit method is one of the candidate candidates, which is not restricted CFL condition. Therefore, ADI-FDTD method is known unconditional stability that it has stability and time step can be large than FDTD method[2].

In this paper, it is confirmed the validity and usefulness of the ADI-FDTD method. As examples, the propagation in free space and the scattering by a circular cylinder is applied [3][4]. The first order Mur absorbing boundary condition is used.

2 ADI-FDTD Method in Two Dimensional TM Wave

In this paper, the E-polarized two-dimensional problem is considered. There are three components of E_z, H_x, H_y. For this case, the algorithm of ADI-FDTD is expressed as follows: [5].

$(The\ first\ step)$

$$\begin{aligned} E_z^{n+1/2}(i,j) = E_z^n(i,j) + J_n + \left(\frac{\Delta t}{2\varepsilon \Delta x}\right)\left\{H_y^{n+1/2}(i+1/2,j) - H_y^{n+1/2}(i-1/2,j)\right\} \\ - \left(\frac{\Delta t}{2\varepsilon \Delta y}\right)\left\{H_x^{n}(i,j+1/2) - H_x^{n}(i,j-1/2)\right\} \end{aligned} \tag{1}$$

$$H_x^{n+1/2}(i,j+1/2) = H_x^n(i,j+1/2) - \left(\frac{\Delta t}{2\mu \Delta y}\right) \cdot \{E_z^n(i,j+1) - E_z^n(i,j)\Big\} \tag{2}$$

$$H_y^{n+1/2}(i+1/2,j) = H_y^n(i+1/2,j) + \left(\frac{\Delta t}{2\mu \Delta x}\right) \cdot \{E_z^{n+1/2}(i+1,j) - E_z^{n+1/2}(i,j)\Big\} \tag{3}$$

$(The\ second\ step)$

$$\begin{aligned} E_z^{n+1}(i,j) = E_z^{n+1/2}(i,j) + J_n + \left(\frac{\Delta t}{2\varepsilon \Delta x}\right)\left\{H_y^{n+1/2}(i+1/2,j) - H_y^{n+1/2}(i-1/2,j)\right\} \\ - \left(\frac{\Delta t}{2\varepsilon \Delta y}\right)\left\{H_x^{n+1}(i,j+1/2) - H_x^{n+1}(i,j-1/2)\right\} \end{aligned} \tag{4}$$

$$H_x^{n+1}(i,j+1/2) = H_x^{n+1/2}(i,j+1/2) - \left(\frac{\Delta t}{2\mu \Delta y}\right) \cdot \{E_z^{n+1}(i,j+1) - E_z^{n+1}(i,j)\Big\} \tag{5}$$

$$H_y^{n+1}(i+1/2,j) = H_y^{n+1/2}(i+1/2,j) + \left(\frac{\Delta t}{2\mu \Delta y}\right) \cdot \{E_z^{n+1/2}(i+1,j) - E_z^{n+1/2}(i,j)\Big\} \tag{6}$$

(i, j) is the index number of the cell in the x and y axes and n is the time step. The expression of the first step is (1)-(3), and the expression of the second step is (4)-(6). In addition, Eq.(1) defines both the left-hand side of $E_z^{n+1/2}$ and the right-hand side of $H_y^{n+1/2}$ same time, Eq.(3) defines both the left-hand side of $H_y^{n+1/2}$ and the right-hand side of $E_z^{n+1/2}$ same time. Similarly, Eq.(4) defines both the left-hand side of E_z^{n+1} and the right-hand side of H_x^{n+1} same time, Eq.(6) defines both the left-hand side of H_x^{n+1} and the right-hand side of E_z^{n+1} same time. We can not calculate to contain different number of $E_z^{n+1/2}$ and $H_y^{n+1/2}$ in Eqs.(1) and (3) due to contained the same time step. After substitute the right-hand side of Eq.(3) for $H_y^{n+1/2}$ in Eq.(1), the Crank-Nicholson type expression can be obtained as follows:

$$\begin{aligned}
&-\left(\frac{\Delta t}{2\sqrt{\mu\varepsilon}\Delta x}\right)^2 E_z^{n+1/2}(i+1, j) + \left\{1 + 2\left(\frac{\Delta t}{2\sqrt{\mu\varepsilon}\Delta x}\right)^2\right\} \cdot \{E_z^{n+1/2}(i, j)\} \\
&\qquad\qquad - \left(\frac{\Delta t}{2\sqrt{\mu\varepsilon}\Delta x}\right)^2 E_z^{n+1/2}(i-1, j) \\
&= E_z^n(i, j) + J_n + \left(\frac{\Delta t}{2\varepsilon\Delta x}\right)\left\{H_y^n(i+1/2, j) - H_y^n(i-1/2, j)\right\} \\
&\qquad\qquad - \left(\frac{\Delta t}{2\varepsilon\Delta y}\right)\left\{H_x^n(i, j+1/2) - H_x^n(i, j-1/2)\right\}
\end{aligned} \tag{7}$$

$$\begin{aligned}
&-\left(\frac{\Delta t}{2\sqrt{\mu\varepsilon}\Delta y}\right)^2 E_z^{n+1}(i, j+1) + \left\{1 + 2\left(\frac{\Delta t}{2\sqrt{\mu\varepsilon}\Delta y}\right)^2\right\} \cdot \{E_z^{n+1}(i, j)\} \\
&\qquad\qquad - \left(\frac{\Delta t}{2\sqrt{\mu\varepsilon}\Delta y}\right)^2 E_z^{n+1}(i, j-1) \\
&= E_z^{n+1/2}(i, j) + J_n + \left(\frac{\Delta t}{2\varepsilon\Delta x}\right)\left\{H_y^{n+1/2}(i+1/2, j) - H_y^{n+1/2}(i-1/2, j)\right\} \\
&\qquad\qquad - \left(\frac{\Delta t}{2\varepsilon\Delta y}\right) \cdot \left\{H_x^{n+1/2}(i, j+1/2) - H_x^{n+1/2}(i, j+1/2)\right\}
\end{aligned} \tag{8}$$

In particular, Eqs.(7) and (8) are the simultaneous linear equations and are solved by using algorithm of Thomas for tridiagonal matrix of structural matrix.

3 Numerical Results

In this section, it is confirmed the validity of the ADI-FDTD method using the algorithms mentioned in the previous section.

3.1 Propagation in Free Space

Figure 1 shows the geometry of the propagation problem in free space. The source of the incident is the line source and the distance between the source and the observation point A is 11.6[μm].

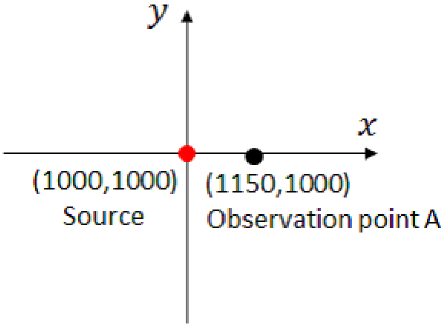


Fig. 1 Geometry of the propagation problem in free space(model 1)

The parameter of the model 1 as shown in Fig. 1 set the wavelength $\lambda = 1.55 \times 10^{-6}$[m], cells size $\Delta x = \Delta y = 7.75 \times 10^{-8}$[m] and number of the cells 2000×2000, which correspond to 15.5 × 15.5[μm]. Incident wave use cylindrical wave of Eqs.(9).

$$J_n = \sin(2\pi f(nt + 1/2)) \tag{9}$$

We set $\Delta t = 1.65 \times 10^{-16}$[s] of the FDTD method by CFL condition. In addition, we set $\Delta t = 1.65 \times 10^{-16}$[s] and $\Delta t = 8.25 \times 10^{-16}$[s] of the ADI method. Time step $\Delta t = 8.25 \times 10^{-16}$[s] is five times greater than $\Delta t = 1.65 \times 10^{-16}$[s]. We define $\Delta t_1 = 1.65 \times 10^{-16}$[s]. Also, we define $\Delta t_2 = 8.25 \times 10^{-16}$[s]. We will calculate using these time step.

Figure 2 shows the normalized amplitude E_z for the different time step. The red and blue lines shows the amplitude of the propagation wave for Δt_1 and Δt_2, respectively. For the large time step, the normalized amplitudes are deviated from each other. This deviation is caused by the numerical dispersion. In addition, it is considered that the large time step can not follow the incident wave in the time domain.

Table 1 shows the calculation time of the conventional FDTD method and the ADI-FDTD method with the different time step for the propagation problem in free space.

Figure 3 shows propagation wave along the x axis. Although ADI method have two step, ADI-FDTD method is faster than the conventional FDTD method for the large time step. It is noted that the time step have to be selected to follow the incident wave.

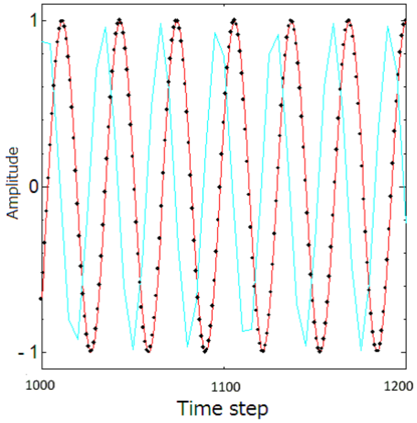


Fig. 2 Waveform at observation point A in time domain
(Black point: FDTD method, Red point and Blue lines: ADI-FDTD method)

Table 1 Comparison of Calculation time

analysis method	calculation time[s]
FDTD method	211
ADI method(Time stepΔt_1)	385
ADI method(Time stepΔt_2)	79.2

3.2 Scattering Problem

In this subsection, the two-dimensional scattering problem is considered and the validity of ADI-FDTD method is examined.

Figure 4 shows the geometry of the two-dimensional scattering problem. The scatterer is assumed to be a circular cylinder.We observe time response at observation point B.

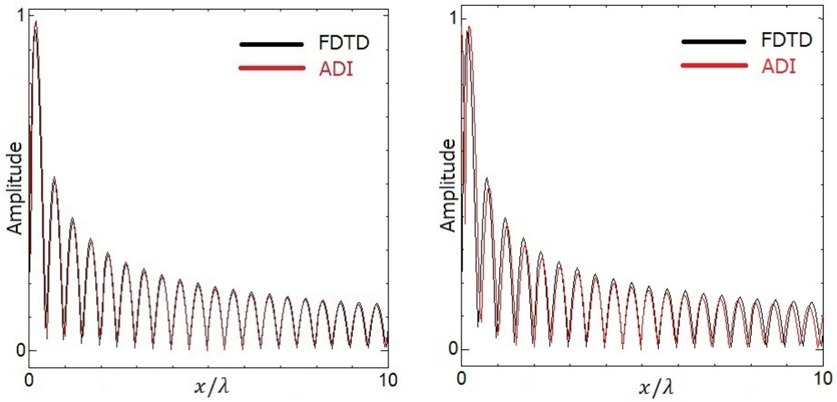
Parameters of the model 2 is the same parameters as that of model 1. The scatterer is the circular cylinder with the relative permittivity $\varepsilon_r = 30$ and diameter of $3.875[\mu m]$. Incident wave is assumed to be the the cylindrical wave of Eq.(9).

Figures 5 and 6 show the contour map of the electric field using the conventional FDTD method and the ADI-FDTD method, respectively.

By comparing the results of Figs. 5 and 6, it is found that the electric field by the ADI-FDTD method is smaller than that by the conventional FDTD method. This is due the same reason of the propagation problem.

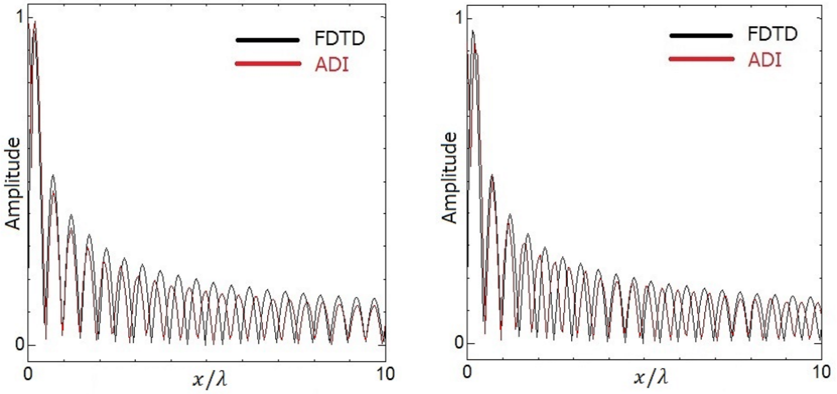
Figure 7 shows the normalized amplitude of the electric filed in the time domain at the observation point B. The red and blue points correspond to the electric field for the time step Δt_1 and Δt_2, respectively.

Figure 8 shows the electric filed which is escalated part of Fig. 7.



(a) Case of Δt_1.

(b) Case of 2 times of Δt_1.



(b) Case of 4 times of Δt_1.

(b) Case of 6 times of Δt_1.

Fig. 3 Normalized amplitude E_z for the different time step size

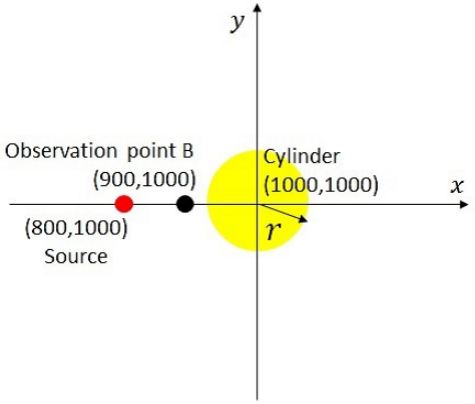

y
Observation point B
(900,1000)
Cylinder
(1000,1000)
x
(800,1000)
Source
r

Fig. 4 Geometry of the scattering problem (model 2)

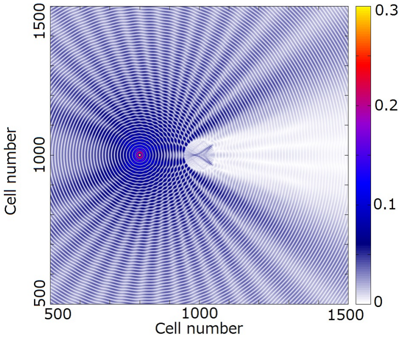
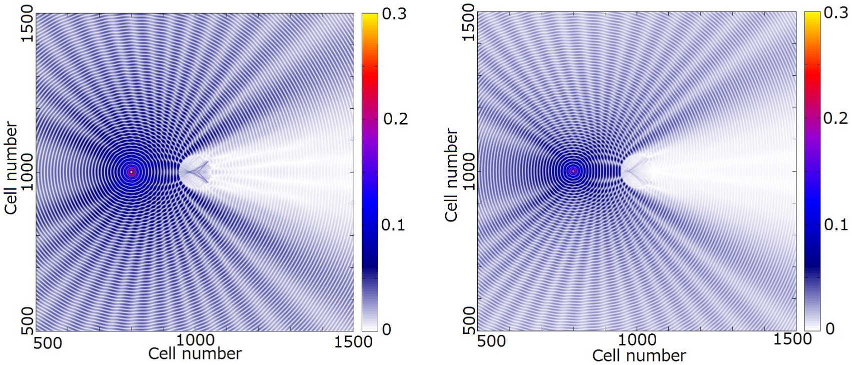


Fig. 5 Contour map of electric field for analytical model 2 by using the conventional FDTD method



(a) Case of Δt_1.

(b) Case of Δt_2.

Fig. 6 Contour map of electric field for analytical model 2 by using the conventional ADI-FDTD method.

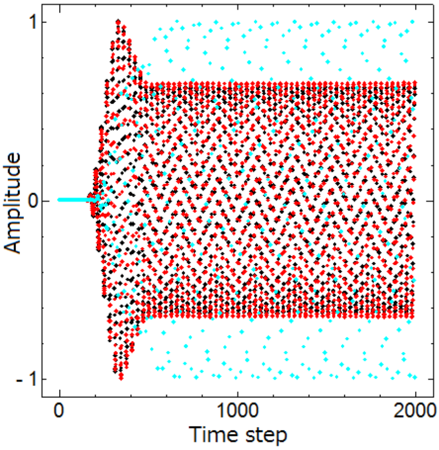


Fig. 7 time waveform of E_z
(Black point: FDTD method, Red point, Blue point: ADI-FDTD method)

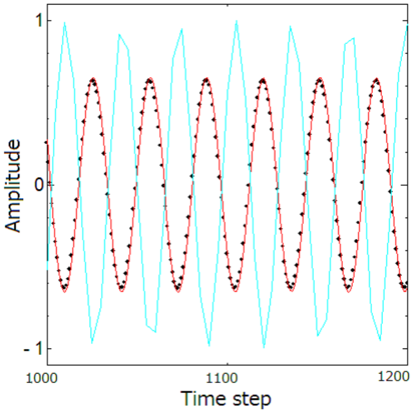


Fig. 8 time waveform of E_z
(Black point: FDTD method ARed line ABlue line: ADI-FDTD method)

It is seen that the blue point can't be captured incident wave, and it shows wrong wave form in Figs. 7 and 8. The black point by the conventional FDTD method is the same as that by the ADI method, which are shown in the red point and follows the incident wave.

4 Summary

The validity and usefulness of the ADI-FDTD method has been examined using the propagation problem in free space and the scattering problem.

From the numerical examinations, it has shown that the ADI-FDTD method doesn't diverge if it use large time step. However, if the time step increase 5 times the validity and the usefulness decreases, since the incident wave with large time step can not follow the original incident wave form. Also, the phase of the ADI method is faster than that of the conventional FDTD method in case of free-space propagation, so we can't get same data for FDTD method.

Due to the above consideration, the ADI-FDTD method is limited and it is noted that the time step have to be selected to follow the incident wave. The cell size in the space domain is also taken carefully. The multigrid method is one of the useful approaches. The combination of the ADI and the multigrid method is the future work.

References

1. Taflove, A.: Computational Electrodynamics: The Finite-Difference Time-Domain Method. Artech House, Norwood (1995)
2. Namiki, T.: A New FDTD Algorithm Based on Alternating-Direction Implicit Method. IEEE Trans. on Microwave Theory Tech. **47**(10), 2003–2007 (1999)
3. García, S.G., Lee, T.-W., Hagness, S.C.: On the accuracy of the ADI-FDTD method. IEEE Antennas Wireless Propag. Lett. **1**(1), 31–34 (2002)
4. Wang, S., Teixeira, F.L., Chen, J.: An iterative ADI-FDTD with reduced splitting error. IEEE Microwave Wireless Comp. Lett. **15**(2), 92–94 (2005)
5. Namiki, T., Ito, K.: Investigation of Numerical Errors of the Two-Dimensional ADI FDTD Method. IEEE Trans. on Microwave Theory Tech. **48**(11), 1950–1956 (2000)

Numerical Examination on Propagation Loss Properties in Case Human Bodies Exist Near the Transmitter Using the FDTD Method

Shota Sakakibara and Mitsuhiro Yokota

Abstract In mobile communications, the communication under various environments is increasing by the rapid spread of the number of the portable terminals. Therefore, many examinations have been reported about the outdoor propagation loss characteristic, and the propagation model and the presuming method are standardised. In this paper, a human body is assumed to be a loss dielectric cylinder, and the effect of the human bodies exist around the transmitter on the propagation loss is examined numerically by using FDTD method.

Keywords Numerical techniques · FDTD method · Propagation loss

1 Introduction

Many more customers are now enjoying various wireless communication serviced such as cellular phone and WLAN, etc. in door environments. In this case, the mobile terminals (MTs) frequently remain stationary while in use. As a result, the communication characteristics are wildly different from those seen in outdoor mobile use for example. Therefore, it is very important to model the fading characteristics in above environment.

In static usage, the MT itself doesn't move, but the environment around it changes due to moving obstacles such as people. In order to accurately evaluate the resulting propagation environment, a propagation model (channel model) that can simulate this environment is necessary. In this case, mobile terminals used in a stationary condition receive a level variation different from that in a level variation different from that a physical channel model for a static terminal used in indoors [1]. The physical channel

---

S. Sakakibara · M. Yokota(✉)
Department of Electrical and Electronic Engineering, University of Miyazaki,
1-1 Gakuen Kibanadai Nishi, Miyazaki 889-2192, Japan
e-mail: hc11036@student.miyazaki-u.ac.jp, m.yokota@m.ieice.org

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/978-3-319-23204-1_42

model for static terminals used indoors was proposed. The proposed model can be considered physical parameters such as moving people, their moving speed and etc. in order to evaluate various situations preciously. An experimental verification [2] has been performed.

In this paper, the effect of the human bodies exist around the transmitter on the propagation loss in free space is examined numerically when it is assumed that the transmitter and the receiver is fixed. As the numerical method, FDTD method is used [4], which transforms Maxwell's equation to the form of differential equations and is a method to solve in the time domain whose basic algorithm has been proposed by K.S.Yee [4]. A human body is assumed to be a loss dielectric cylinder that has relative permittivity ε_r and permeability μ and conductivity σ. The electromagnetic wave scattering by loss dielectric cylinders is examined in order to estimate the propagation loss. Mur's absorbing boundary condition of second order is applied as absorbing boundary conditions.

2 Numerical Analysis

We describe setting of numerical analysis using the FDTD method. The computational domain of the FDTD method is shown in Fig. 1. The dot and dashed line means the basic unit area, which is defined 2[m] × 2[m] in this paper, and the loss cylinder which assumed that human bodies are placed in this basic area.Each basic unit is indicated by Group 1, 2, and 3 in Fig. 1, respectively. Around the transmitter, we set the entry-impossible area with a radius of 0.35[m] where human bodies are not placed.

Also, as shown Fig. 2, it is divided the Group 1 into two parts in order to see the effects around the transmitter.

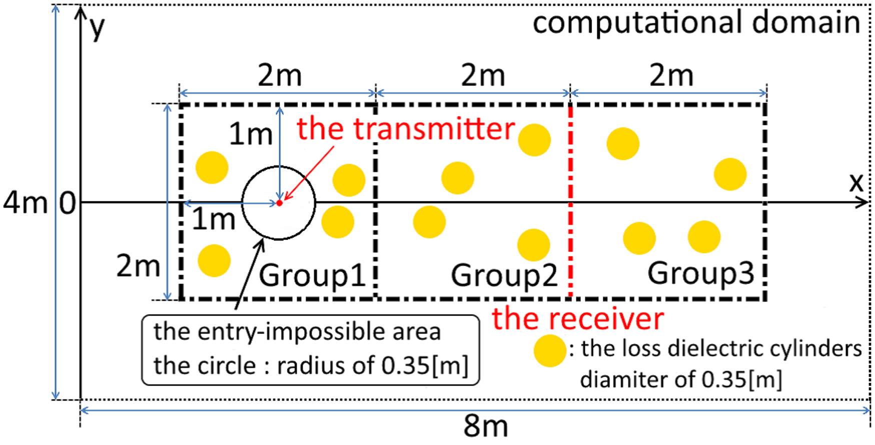

y
computational domain
2m
2m
2m
1m
the transmitter
x
4m
0
1m
2m
Group1
Group2
Group3
the receiver
the entry-impossible area
the circle : radius of 0.35[m]
: the loss dielectric cylinders
diamiter of 0.35[m]
8m

Fig. 1 Geometry of the problem(Case 1)

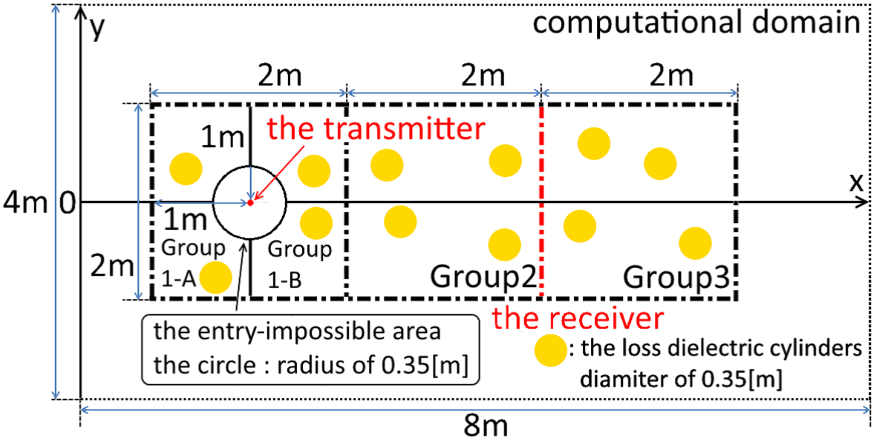

y
computational domain
2m
2m
2m
1m
the transmitter
x
4m
0
1m
Group 1-A
Group 1-B
2m
Group2
Group3
the receiver
the entry-impossible area
the circle : radius of 0.35[m]
: the loss dielectric cylinders
diamiter of 0.35[m]
8m

Fig. 2 Geometry of the problem(Case 2)

The center of the loss dielectric cylinders are arranged between 1[m]$\leq$x$\leq$ 3[m], 3[m]$\leq$x$\leq$ 5[m], 5[m]$\leq$x$\leq$ 7[m] and -1[m]$\leq$y$\leq$ 1[m],at each group randomly and the trial number is chosen as 50.

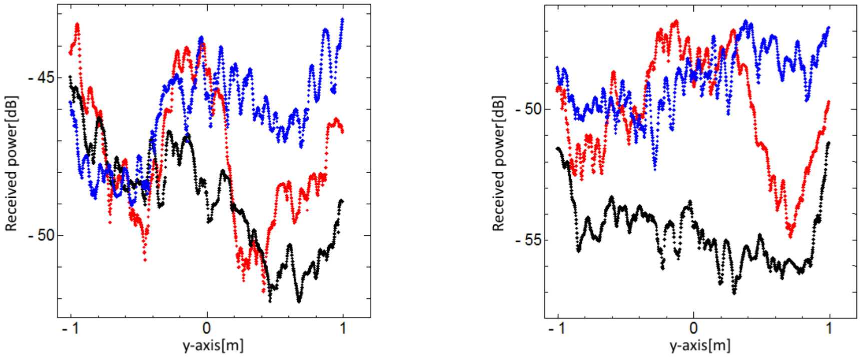
As the parameters, the frequency $f = 3.35$[GHz], The number of cell 4000×2000, size of cell Δx= Δy= 2×10^{-3}[m], relative permittivity $\varepsilon_r = 50$, and conductivity $\sigma = 2$ [S/m] are assumed [5], respectively. The diameter of loss dielectric cylinder is chosen by 0.35[m], which is obtained by the proposed model in [1], [2]. The number of calculations of the time is $nt = 10000$ times.

The incident wave is given by Eq. (1), which corresponds to a line source of wave and a continuous wave.

$$E_z^n(i, j) = \sin(2\pi f n \Delta t) \tag{1}$$

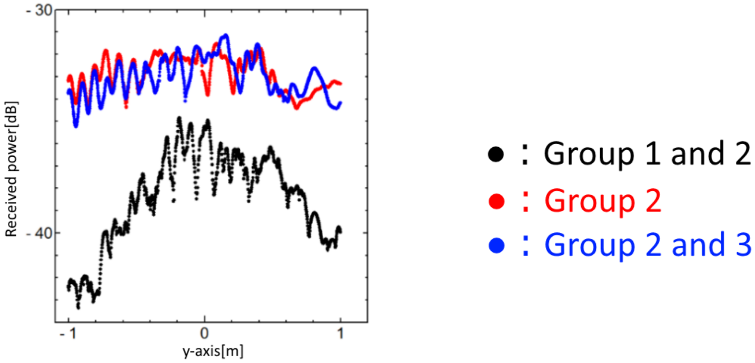
2.1 In a case of Changing the Placement Pattern of Human Bodies and the Number of Human Bodies

In this section, the effect of the placement pattern of human bodies on the propagation loss is examined. In order to the effect of the placement pattern of a human body on the propagation loss is examined, the averaged received powers are calculated for various kinds of combination of the groups. Human bodies are placed one of the case of 4 cylinders on the other hand the case of 6 cylinders and the case of 8 cylinders in each group. The averaged powers are defined in this paper by averaging the received power at the observation point for each trial. Figure 3 shows the averaged received power in order to examine the effect of the scattering from the backward. Table 1 shows the averaged received power for three kinds of the number of the cylinders for three kinds of the placement patterns. The averaged received powers in Table 1 are defined by the averaging the received power at the observation point between -1[m]$\leq$y$\leq$ 1[m] in Fig. 1.



(a) Case of 4 cylinders.

(b) Case of 6 cylinders.



(c) Case of 8 cylinders.

Fig. 3 An average received power distribution for three kinds of the placement patterns of three kinds of the number of cylinders.The black, red, and blue marks indicate the Group1 and 2, Group 2, and Group 2 and 3, respectively.

Table 1 An average received power in each group

(a) Case of 4 cylinders.

Group	average received power[dB]
Group 1,2	-48.93
Group 2	-47.60
Group 2,3	-46.33

(b) Case of 6 cylinders.

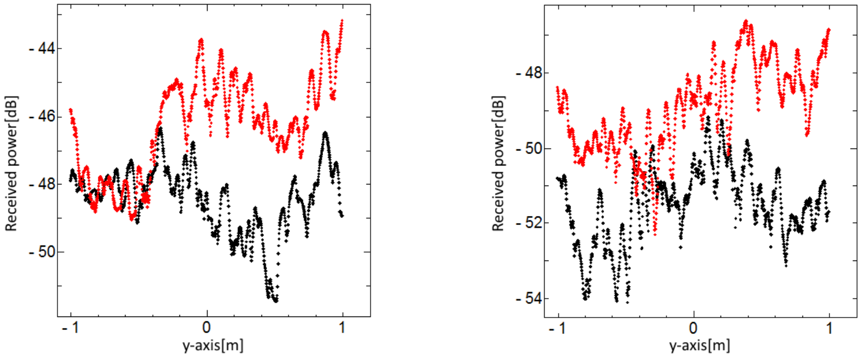
Group	average received power[dB]
Group 1,2	-54.84
Group 2	-50.16
Group 2,3	-48.96

(c) Case of 8 cylinders.

Group	average received power[dB]
Group 1,2	-57.77
Group 2	-55.45
Group 2,3	-52.32

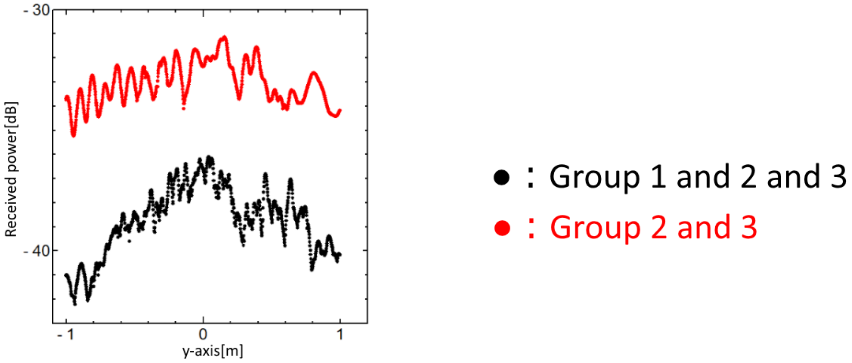
When 4 human bodies is placed in each group, the averaged power becomes 1.33[dB] larger than the other cases due to the scattering from the Group 3 from Table 1 (a). Similarly, the averaged power for the case of 6 and 8 cylinders becomes 4.68[dB] and 2.32[dB] larger than the other cases. The tendency such as the averaged received power becomes larger due to the scattering from the backward of the observation point is confirmed by the experiment.

Figure 4 shows the averaged received power in order to examine the effect of the human bodies around the transmitter.



(a) Case of 4 cylinders.

(b) Case of 6 cylinders.



(c) Case of 8 cylinders.

Fig. 4 An average received power distribution for two kinds of the placement patterns for two kinds of three kinds of the number of cylinders. The black and red marks indicate the Group 1, 2, 3, and Group 2, 3, respectively.

When 4 human bodies is placed in Group 2, 3, the averaged power becomes 2.18[dB] larger than the other cases. Similarly, the averaged power for the case of 6 and 8 cylinders becomes 6.74[dB] and 3.82[dB] larger than the other cases.

Table 2 An average received power in each group

(a) Case of 4 cylinders.

Group	average received power[dB]
Group 1,2,3	-48.51
Group 2,3	-46.33

(b) Case of 6 cylinders.

Group	average received power[dB]
Group 1,2,3	-55.70
Group 2,3	-48.96

(c) Case of 8 cylinders.

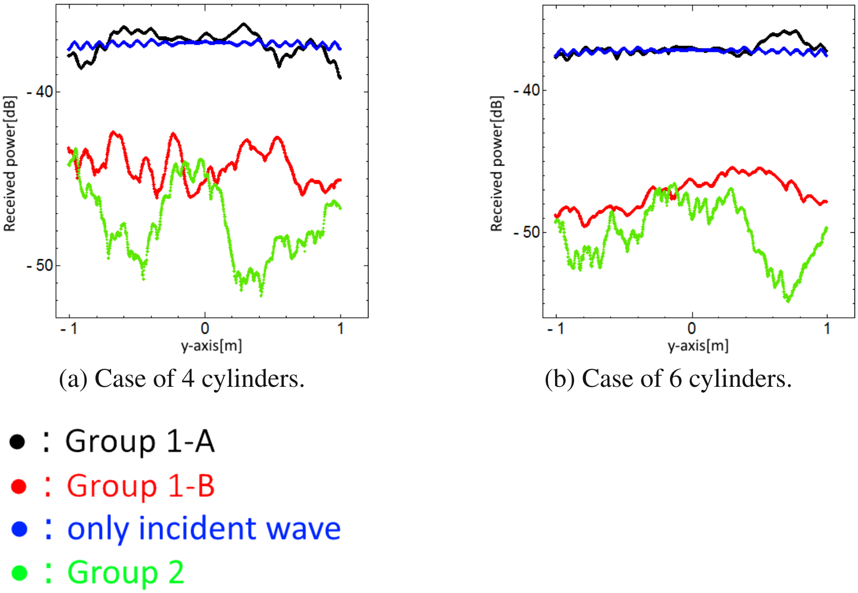
Group	average received power[dB]
Group 1,2,3	-56.14
Group 2,3	-52.32

From Fig. 4 and Table 2, it is found that the averaged power is decreased due to the attenuation by the human bodies which is placed in Group 1.

2.2 In a Case of Human Bodies are Placed Only Around the Transmitter

Group 1 is divided Group 1-A of backward the transmitter and Group 1-B of forward the transmitter, as shown in Fig. 2, and the propagation loss is examined in order to examine the effect of the human bodies around the transmitter. Figure 5 shows the averaged received power in case of human bodies is placed four or six each in each group. Also, Table 3 shows the averaged received power in each group.

When the human bodies are placed at the backward of the transmitter, the averaged received power at the observation point is the same as that of the incident wave. This means that the human bodies backward of the transmitter does not affect the propagation loss. On the other hand, the averaged received power for Group 1-B has almost the same property as that of Group 2. It is found that the human bodies are placed in front of the transmitter and human bodies affect the propagation loss. It is seen from Table 3 that the difference of the averaged power for 4 and 6 human bodies between the Group 1-B and Group 2 are, respectively, 3.24[dB] and 2.84[dB].



(a) Case of 4 cylinders. (b) Case of 6 cylinders.

Fig. 5 An average received power distribution around the transmitter for two kinds of the placement of the cylinder around transmitter. The black and red lines mean the received power for Group 1-A and Group 1-B, respectively. The blue line indicates the incident wave distribution, and the green line means the received power for Group 2.

Table 3 An average received power in each group

(a) Case of 4 cylinders.

Group	average received power[dB]
Group 1-A	-37.29
Group 1-B	-44.36
only incident wave	-37.33
Group 2	-47.60

(b) Case of 6 cylinders.

Group	average received power[dB]
Group 1-A	-37.13
Group 1-B	-47.32
only incident wave	-37.33
Group 2	-50.16

As the passerby density of Group 1-B is twice of the original configuration, the averaged received power distribution is similar as that of Group 2 and approaches to that of Group 2 as the passerby density increases. propagation

3 Conclusions

In this paper, the effect of the human bodies exist around the transmitter on the propagation loss has been examined numerically by using FDTD method. We confirmed that the received power increases when the human bodies exist the backward the receiver. Also, we confirmed that the received power decreases when the human bodies exist around the transmitter. The effect of the propagation loss on the placement for the backward the transmitter has not been confirmed, because we set 3[m] between the transmitter and the receiver. So that we needs to be examined when the transmitter exist near the receiver as the future works.

References

1. Fujii, T., Ohta, Y.: Proposal of Fading Model in Indoor Environment (1) -In case of Fixed Terminal and NLOS-. IEICE Tech. Report, AP2005-55 (2007)
2. Ohta, Y., Fujii, T.: An Empirical Study on Radio Wave Shadowing by Human Body. IEICE Tech. Report, AP2008-159 (2009)
3. Uno, T.: Electromagnetic field and antenna analysis by FDTD method. Corona Publishing, Tokyo (1998)
4. Taflove, A.: Computational Electrodynamics: The Finite-Difference Time-Domain Method. Artech House, Norwood (1995)
5. Johnson, C.C., Guy, A.W.: Nonionizing electromagnetic wave effects in biological materials and systems. Proc. IEEE **60**, 692–718 (1972)
6. Yokota, M., Shiiya, K., Ohta, Y., Fulii, T.: Propagation loss properties in case human bodies exist between transmitter and receiver. In: Proc. of 2012 ISAP, pp. 343–346 (2012)

Measurement of Udder Surface Temperature in Cows Using Infrared Thermometer

Yuki Nakagawa, Noor Agha Nassary, Kiichi Fukuyama and Ikuo Kobayashi

Abstract A total of 28 cows in various stages of lactation were used in this study. Measurement of surface of the udder temperature was carried out using infrared thermometer (IRT), totalizing 400 measurements. Differences of temperature between animals infected and uninfected with mastitis, breeds, number of parturition and hind quarters of the udder were examined. Surface temperature of the left part of the udder was significantly higher than that of the right part of the udder in dairy cows. Udder temperature in Jersey cows was significantly higher than that of Holstein cows. The udder temperature at 9:00 was lower than the others. IRT equipment is easy to handle and quick to measure the body temperature of animals in the field. Future development of ICT system such as alerting the producer potential problem animals, by using automated IRT scanning mechanism, would be a helpful and effective method to improve livestock productivity.

Keywords Dairy cows · Infrared thermometer · Mastitis · Body surface temperature

1 Introduction

Body temperature is an important indicator in the diagnosis of mammals' illnesses and for estimation of their physiological status. There are many methods to measure body temperature; however, non-contact and non-invasive infrared measurement is often the better method for the animals in the field. It would be valuable to develop new screening methods for the identification of animals infected with diseases without catching and restraining of affected animals.

Y.Nakagawa · N.A. Nassary · K. Fukuyama · I. Kobayashi(✉)
Graduate School of Agriculture, University of Miyazaki, Miyazaki, Japan
e-mail: ikuokob@cc.miyazaki-u.ac.jp

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/ 978-3-319-23204-1_43

Mastitis is one of the most serious diseases causing economic loss in dairy farms. It is a major endemic disease of dairy cattle, and is the inflammation of the mammary gland and udder tissue. It usually occurs as an immune response to bacterial invasion of the teat canal by variety of bacterial sources present on the farm. The most obvious symptoms of clinical mastitis are abnormalities in the udder such as swelling, heat, hardness, redness, or pain; and the milk has a watery appearance, flakes, clots, or pus. Other symptoms can also include a reduction in milk yield, the lack of appetite, and dehydration, especially an increase in body temperature.

Infrared thermography has potential application in the clinical monitoring inflammation [1] and for assessment of animal welfare [2]. Infrared thermography measures the radiant energy from the surface of a source, which can be used to calculate its surface temperature. Contactless surface temperature measurement using an infrared thermometer (IRT) could be a modern, safe technique of thermal profile visualization to detect livestock diseases such as mastitis. Also, this technology may be useful for assessing the physiological responses of dairy cows to milking and feeding. Future development of ICT system such as alerting the producer of potential problem animals, by using automated IRT scanning mechanism, would be a helpful and effective method to improve livestock productivity.

Consequently, the aim of this preliminary study was to investigate the relationship between mastitis and the skin surface temperature of cow's udder using low-cost hand held infrared thermometers, in order to establish framework for detecting abnormalities in the cows.

2 Materials and Methods

A total of 28 cows (Holstein Friesian: n = 17, Jersey cows: n = 3, Japanese Black cows: n = 8, parity 2.8 ± 0.4, mean ± SEM) in various stages of lactation were used in this study. Animals raised on Sumiyoshi Livestock field station (total area = 50 ha), Faculty of Agriculture, university of Miyazaki, were enrolled in the experiment in May, 2015. These cows were pastured from 09.00 – 15.00 hours on 0.5 – 3.2 ha of pasture (Italian grass, Guinea grass and Rhodes grass), depending on the quantity of available grass, and kept in a freestall barn from 15.00 – 09.00 hours. Additional hay, wrapped silage, corn silage and concentrates were provided every afternoon at 15.00 hours. Lactating cows were milked twice daily at milking parlor. Morning milking commenced at 08.30 and afternoon milking at 17.30. Milk samples from each quarter for each cow were obtained, and samples were analyzed for somatic cell count (SCC) using cell counter (DeLaval). Definition of clinical mastitis was based on SCC at the beginning of the study which was over 500,000.

Images of the posterior surface of the left and right hind quarters of the udder were taken directly behind the standing animals. The image was taken at a distance of 2m. Measurement of surface udder temperature was carried out using IRT (SANSYO, SIK-300), totalizing 400 measurements. Data of environmental temperature was obtained from Japanese Meteorological Agency to estimate the effect on body surface temperature of the animals. We examined the relationship of udder surface temperature and environmental temperature, differences of temperature between animals infected and uninfected with mastitis, breeds, number of parturition and hind quarters of the udder. Mann–Whitney *U*-test and Tukey–Kramer test were used to determine differences in the groups. All of the protocols were approved by the Institutional Review Board for animal experiments of the University of Miyazaki.

3 Results and Discussions

Fifty-five measurements out of 400 were carried out from the cows infected with mastitis. There was no difference between normal animals and infected animals with mastitis. There was a difference of temperature between posterior surface of the left and right hind quarters of the udder. Surface temperature of the left part of the udder was significantly higher than that of the right part of the udder in multiparous Holstein cows (Fig. 1.). There was no difference between left and right udder temperature in primiparous Holstein cows, Jersey cows and Japanese Black cows. Temperature of the left part of the udder in Jersey cows was significantly higher than that of Holstein cows. There was a significant difference between primiparous Holstein cows and Jersey cows in the right part of the udder. Correlations between udder surface temperature and environmental temperature were not observed.

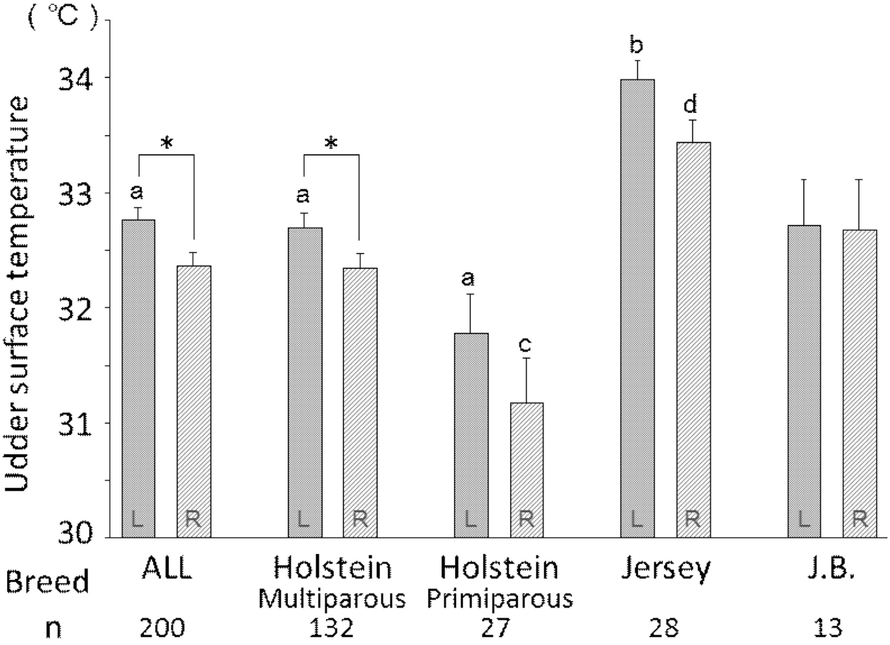


Fig. 1 Differences in temperature between left and right part of the udder surface in different breeds (ab < 0.05, cd < 0.05)

In this study, there was a significant difference of temperature between posterior surface of the left and right hind quarters of the udder. The degree of difference in human thermal asymmetry between opposite sides of the body is generally believed to be subtle, because blood flow governed by the autonomic nerve impulse seems to be symmetrical, both anatomically and histologically. Asymmetric temperature changes on the body surface may be diagnosed as diseases relating to autonomic functions [3]. We could not confirm the reason why temperature of left udder was significantly higher than that of right udder in this study. Although it was reported that there was no significant differences between temperature of left and right udder quarters before and after milking in the cows measured by thermography [4], there may be some different heat radiation system in the body of dairy cows because distribution of stomach and intestines is asymmetrical in the ruminants. Ruminants have a big rumen which produces much heat by rumination in the left side of the body, and it may effect on temperature of the surface skin of udder in lactating dairy cows.

There was a difference of udder surface temperature depending on the time of measurement (Fig. 2.). The udder surface temperature at 9:00 was significantly lower than that at 14:00, 17:00 and 20:00. Many studies have shown that cattle have a circadian rhythm in body temperature [5,6]. Measurement of udder surface temperature to detect abnormalities would need to account for daily circadian rhythm and various environmental conditions. Detailed monitoring of environmental temperature may also improve accuracy for detecting changes in body surface temperature of animals.

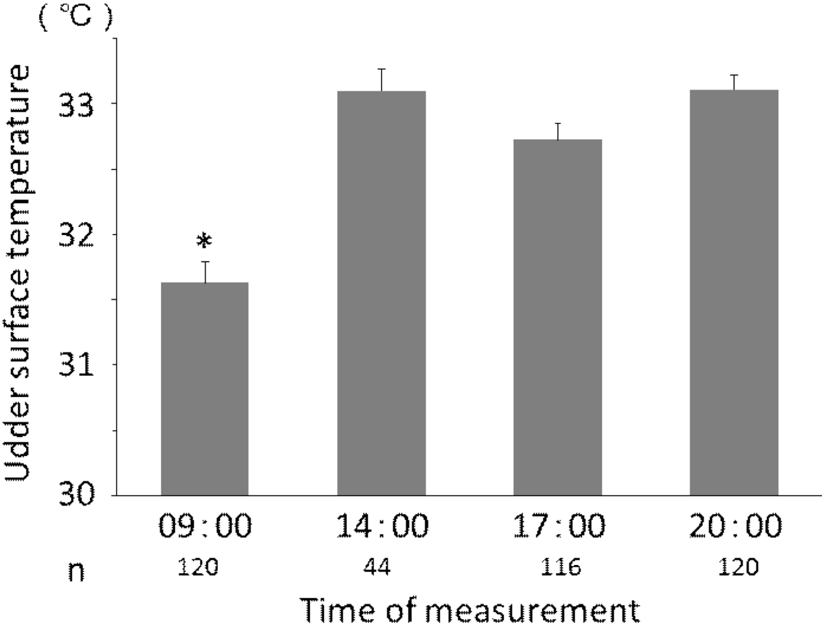


Fig. 2 Relationship between udder surface temperature and measurement time

Thermographic equipment has found increasing applicability, as infrared thermography is a non-invasive safe method and only minimal restraint may be necessary. Although we could not find difference of udder temperature between normal cows and cows infected with mastitis in the study, IRT has been used to predict changes in udder temperature for early diagnosis of mastitis in dairy cows [7,8,9]. Increase in body temperature of cows with experimentally induced clinical mastitis was detected by thermal camera [10]. Some researchers found that IRT may be useful for early diagnosis of laminitis [11], and a significant difference in temperature of the coronary band between cows with lesions and cows without lesions was reported [12]. It has also been reported that IRT have been tested for early detection of foot-and-mouth disease virus infected cattle [13].

Some researchers found that infrared thermography has potential application for assessment of animal welfare [2]. One of the concerns regarding animal welfare is heat stress which is a serious problem for dairy cows in hot season, and it causes mastitis by depressing immune system in the cows. Many of the techniques used to measure stress involve invasive procedures, such as blood sampling, which may cause a stress response. Application for IRT in the management of cattle could be beneficial to improve animal welfare.

In Japan, the average number of animals per livestock farm has increased as the number of holding has declined. The equipment used in this study is easy to handle and quick to measure the body temperature of animals, however, it would be laborious for a producer to measure body temperature for the entire herd on a daily basis. Future development of technology such as automated scanning mechanism, simply alerting the producer potential problem animals, is required. Although recent time advances have been achieved in the use of IRT in studies of animal sciences, there still remain some limitations of this technology, as data should be collected out of direct sunlight and wind drafts, and hair coats of animals should be free of dirt, moisture or foreign material [2]. Also, it is needed to understand the activities of animals, seasonal change of circumstances which may affect body surface temperature, animal physiology including circadian rhythm of the animals and so on. Further studies on the use of IRT are needed before it can be used as a reliable diagnostic tool, in order to establish framework for detecting abnormality in the cows.

References

1. Dennis, D.S., Lene, J.P.: Infrared skin temperature measurements for monitoring health in pigs: a review. Acta Veterinaria Scandinavica **57**, 5 (2015)
2. Stewart, M., Webster, J.R., Schaefer, A.L., Cook, N.J., Scott, S.L.: Infrared thermography as a non-invasive tool to study animal welfare. Animal Welfare **14**, 319–325 (2005)
3. Niu, H.H., Lui, P.W., Hu, J.S., Ting, C.K., Yin, Y.C., Lo, Y.L., Liu, I., Lee, T.Y.: Thermal Symmetry of skin temperature: Normative data of normal subjects in Taiwan. Chinese Medical Journal (Taipei) **64**, 459–468 (2001)

4. Poikalainen, V., Praks, J., Veermae, I., Kokin, E.: Infrared temperature patterns of cow’s body as an indicator for health control at precision cattle farming. Agronomy Research Biosystem Engineering (1), 187–194 (2012)
5. Araki, C.T., Nakamura, R.M., Kam, L.W.G.: Diurnal temperature sensitivity of dairy cattle in a naturally cycling environment. J. Therm. Biol. **12**, 23–26 (1987)
6. Lefcourt, A.M., Huntington, J.B., Akers, R.M., Wood, D.L., Bitman, J.: Circadian and ultradian rhythms of body temperature and peripheral concentrations of insulin and nitrogen in lactating dairy cows. Domest. Anim. Endocrinol. **16**, 41–55 (1999)
7. Berry, R.J., Kennedy, A.D., Scott, S.L., Kyle, B.L., Schaefer, A.L.: Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection. Can. J. Anim. Sci. **83**, 687–693 (2003)
8. Scott, S.L., Schaefer, A.L., Tong, A.K.W., Lacasse, P.: Use of infrared thermography for early detection of mastitis in dairy cows. Can. J. Anim. Sci. **80**, 764–765 (2000)
9. Colak, A., Polat, B., Okumus, Z., Kaya, M., Yanmaz, L.E., Hayili, A.: Early detection of mastitis using infrared thermography in dairy cows. J. Dairy Sci. **91**, 4244–4248 (2008)
10. Hovinen, M., Silvonen, J., Taponen, S., Hänninen, L., Pastell, M., Aisla, A.M., Pyörälä, S.: Detection of clinical mastitis with the help of a thermal camera. J. Daily Sci. **91**, 4592–4598 (2008)
11. Nikkhah, A., Plaizier, J.C., Einarson, M.S., Berry, R.J., Scott, S.L., Kennedy, A.D.: Infrared thermography and visual examination of hooves of dairy cows in two stages of lactation. J. Dairy Sci. **88**, 2749–2753 (2005)
12. Alsaaod, M., Büscher, W.: Detection of hoof lesions using digital infrared thermography in dairy cows. J. Dairy Sci. **95**, 735–742 (2012)
13. Rainwater-Lovett, K., Pacheco, J.M., Packer, C., Rodriguez, L.L.: Detection of foot-and-mouth disease virus infected cattle using infrared thermography. The Veterinary Journal **180**, 317–324 (2009)

River Flood Prediction Using Markov Model

Thinn Htet Htet San and Mie Mie Khin

Abstract With the incidence of severe weather and flooding on the increase around the world, there is a need to improve flood forecasting and warning. Floods cause physical damage, loss of basic sanitation that leads to disease, economic hardship due to rebuilding costs and food shortages. By improving flood forecasts it becomes possible to take mitigating actions in advance of the flood and hence avoid millions of pounds worth of damage and even human fatalities. In this paper, Markov model for river flood prediction is constructed. This model focuses on the prediction of events and can capture the fact that time flows forward. The output will be approximate and show that there is a close agreement between the predicted and actual river flooding amount over Ayeyarwady River in Myanmar.

Keywords Flood · Prediction · Markov · Ayeyarwady river

1 Introduction

A flood is a natural event that can have far reaching effects on people and the environment. A flood is caused by a combination of heavy rainfall causing river / oceans to over flow their banks, and can happen at any time of the year. Floods in Myanmar generally occurs during the south west monsoon season of June to October when the monsoon troughs or low pressure waves superimposed on the general monsoon pattern resulting in intense rainfall over strategic areas of the river catchments. According to recent statistics, temperature in Myanmar have been rising gradually during the past 30 years by about 0.6 Celsius since 1977, resulting in irregular weather patterns, including a shorter but more intense monsoon which

---

T.H.H. San(✉)
University of Computer Studies, Mandalay, Myanmar
e-mail: thinnhtetsan2000@gmail.com

M.M. Khin
Myanmar Institute of Information Technology, Mandalay, Myanmar
e-mail: miemie.khin9@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/ 978-3-319-23204-1_44

had increased the risk of flooding. Floods have affected loss of crops and valuable property and even loss of people. Loss of human life and property etc. can be reduced to a considerable extent by giving reliable advance information about the coming floods.

Flood prediction and control is one of the greatest challenges facing the world today, which have become more frequent and severe due to the effects of global climate change and human alterations of the natural environment. In this work, daily flood level prediction for each weather station is implemented by applying Markov model. It is characterized by real time of data flow, which is from the recording equipment in testing of complex systems, and the rapid processing data, at processing of satellite information. Weather stations are selected by along with Ayeyarwaddy river which often face local flooding in rainy seasons as our region of interest. Flood level can be modeled by using various time scales such as yearly, monthly, weekly and daily.

This paper is organized in the following sequence. A briefly survey of many researchers is presented in section 2. The application area is dealt with in section 3. This is followed by proposed Markov model in section 4. Model evaluation, comparison and error results are discussed in section 5. The system is concluded in section 6.

2 Related Words

U. H. M. Thein and U. S. Tun [11] presented an early flood warning system that drew a plan to moderate floods occurred along the major rivers of Myanmar. They aim to manage all types of disasters including floods using operational procedure of early flood warning which was developed by Department of Meteorology and Hydrology (DMH). They highlighted floods during the peak monsoon period of 2007 in Myanmar.

M. Kannan et al. [3] predicted the rainfall by using multiple linear regression (MLR) models and computed Pearson coefficient for five years data and then compared with predicted data using regression approach. The predicted values lie below computed values. According to the results, it does not show accuracy but show an approximate value. Predicting flood, cyclone, forest fire detection, and global warming was avoided.

M. A. Kulkarni et al. [4] proposed wind speed prediction using four different statistical techniques; Curve fitting, Auto Regressive Integrated Moving Average Model (ARIMA), Extrapolation using periodic curve fitting and Artificial Neural Networks (ANN). They computed the Root Mean Square Error (RMSE) in prediction of zonal component of wind speed for all the months.

In [10] Song et al. presented key findings and methods for modeling and forecasting from 2000-2008, including time-series, econometric, and combination techniques. This survey identified some new research directions, which included improving the forecasting accuracy through integrating both qualitative and quantitative forecasting approaches, tourism cycles, seasonality analysis, impact assessment, and risk forecasting.

N. Sen presented a long-range summer monsoon rainfall forecast model based on power regression technique with the use of Ei Nino, Eurasian snow cover, north west Europe temperature, Europe pressure gradient, 50 hPa Wind pattern, Arabian sea SST, east Asia pressure and south Indian ocean temperature in previous year. The experimental results showed that the model error was 4% in [7].

S. Nkrintra, et al. [6] described the development of a statistical forecasting method for SMR over Thailand using multiple linear regression and local polynomial-based nonparametric approaches. SST, sea level pressure (SLP), wind speed, EiNino Southern Oscillation Index (ENSO), and IOD were chosen as predictors. The experiments indicated that the correlation between observed and forecast rainfall was 0.6. T. Sohn, et al. [9] has developed a prediction model for the occurrence of heavy rain in South Korea using multiple linear and logistics regression, decision tree and artificial neural network. They used 45 synoptic factors generated by the numerical model as potential predictors.

S. Banik et al. (2009) [1] have developed rainfall forecasting model using ANN, ANFIS and GA processes and the results obtained by these models are also compared to the statistical. The ANFIS forecasting model and the GA forecasting model can be used to forecast monthly monsoon rainfall more accurately than the ANN model and the statistical model, forecasting method, namely linear multiple regression model.

D. Nayak et al. (2013) have done survey on rainfall predictions using different neural network architectures over twenty-five years. From the survey it has been found that most of the researchers used back propagation network for rainfall prediction and got significant results. The survey also gives a conclusion that the forecasting techniques that use MLP, BPN, RBFN, SOM and SVM are suitable to predict rainfall than other forecasting techniques such as statistical and numerical methods in [5].

El-Shafie et al. (2011) [8] have developed an adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN) model to forecast the rainfall for Klang River in Malaysia on monthly basis. The result showed performance of ANFIS method is better than ANN method and concluded that ANFIS method is superior to the ANN method in forecasting monthly rainfall.

3 Application Area

Myanmar is located between latitudes 09° 32'N and 28° 31'N and longitudes 92° 10'E and 101° 11'E. The location and topography of the country generate a diversity of climate conditions and seasonal changes in the monsoon wind directions crate summer, rainy and winter seasons. According to historical statistics, heavy monsoon rains in Myanmar especially from mid-July to mid-September have caused flooding along the Ayeyarwady, Chindwin, Thanlwin, Sittaung and Yangon rivers and their tributaries [2].

Ayeyarwady is the country's largest river and most important commercial waterway. It starts in the northern part of the country, Kachin State, and flows south to the Andaman Sea on the Ayeyawady coast. Its drainage area of about 413,000 km^2 covers a larger part of Myanmar. Due to monsoonal rains, which occur between mid-May and mid-October, the volume of the Ayeyarwady and its tributaries varies greatly throughout the year.

The river flows through the cities: Putao, Myitkyina, Bhamaw, Katha, Tagaung, Kyaukmyaung, Mandalay, Sagaing, Chauk, Bagan, Nyaung-U, Magway, Pyay, Hinthada, Pantanaw. Details of some weather stations and geographical area of study in this work are shown in Table 1 and Figure 1, respectively.

In this work, daily flood level prediction for each weather station is implemented by applying Markov model. Using this model we can use evidence of a past days observed values to update the statistical beliefs we have of what might occur in the future days. With these updated statistical beliefs we can then make forecasts.

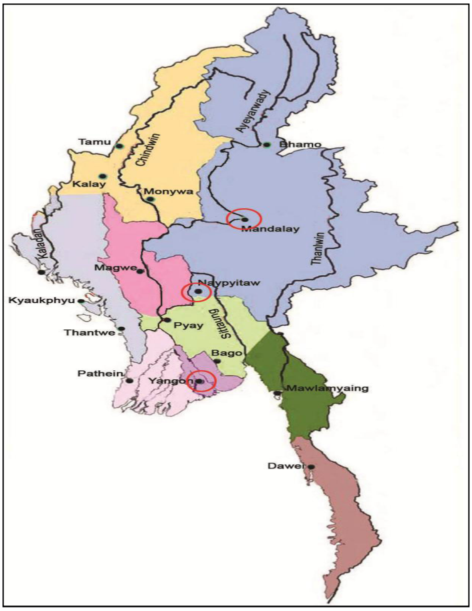

Ayeyarwady
Tamu
Chindwin
Bhamo
Kalay
Monywa
Mandalay
Kaladan
Thanlwin
Magwe
Naypyitaw
Kyaukphyu
Thantwe
Pyay
Sittaung
Bago
Pathein
Yangon
Mawlamyaing
Dawei

Fig. 1 Map of Myanmar with Ayeyarwaddy River

Table 1 Details of Met/Hydro Stations

No.	Station Name	Region	Location
1.	Myitkyinar	Upper	25.3’N, 97.4’E
2.	Sagaing	Central	21.5’N, 95.6’E
3.	Mandalay	Central	21.9’N, 96.1’E
4.	Magway	Central	18.83’N, 93.78’E
5.	Aunglan	Lower	19.38’N, 95.23’E
6.	Pyay	Lower	18.8’N, 95.21’E

The main factors that cause flooding are heavy rainfall, sudden or heavy snow melt, and dam failure, the possibility of levee failure. All of these factors can suddenly increase discharge of water into streams, within streams, and out of streams. Furthermore, when the discharge causes the river to rise above flood stage water runs onto the floodplain.

In this work, spatial dependencies not only between variables but also between weather stations are considered to model the riverine flood using collected hydrological and hydro-meteorological data which are mainly contributing to riverine flood.

4 Markov Chain

A Markov chain is a stochastic process (random process) in which the probability distribution of the current state is conditionally independent of the path of past states, a characteristic called the Markov property. A sequence of states: X_1, X_2, X_3,…(Usually over time). Markov chain is a discrete-time stochastic process with the Markov property. Fig. 2 shows the process of Markov Chain.

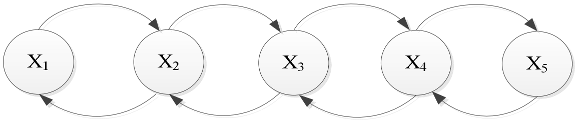


Fig. 2 Markov Chain Process

4.1 Proposed Markov Model

In this section, a model based on the spatial dependencies between variables which are observable factors contributing to the flood level of Ayeyarwady river is build using Markov. The model is used to analyse the current state and movement of a variable to predict the future occurrences and movement of this variable by the use of presently known probabilities. The probability of an event depends on the previous event. Rainfall, Snow melt, Tide, and Water level can be effected river flooding. Among these, the system considered only water level.

To construct the Markov model, Markov states are needed such as Flood (F), Approximately Flood (AF) and Non Flood (NF) using water level. After that the transition probability values are calculated in Table 2. Transition probability matrix and its Markov model for water level at Mandalay station in 1998 is described as shown in Fig. 3 and 4.

Table 2 Probability Transition Values for Mandalay Station during May-Sept in 1998

No.	State	Count	Count / All	Probabilit
1	F_F	7	7 / 153	0.045752
2	F_AF	2	2 / 153	0.013072
3	F_NF	0	0 / 153	0
4	AF_F	2	2 / 153	0.013072
5	AF_AF	41	41 / 153	0.26797
6	AF_NF	2	2 / 153	0.013072
7	NF_F	0	0 / 153	0
8	NF_AF	2	2 / 153	0.013072
9	NF_NF	97	97 / 153	0.63399

$$P = \begin{array}{c} \\ F \\ AF \\ NF \end{array} \begin{array}{c} \begin{array}{ccc} F & AF & NF \end{array} \\ \begin{pmatrix} 0.045752 & 0.013072 & 0 \\ 0.013072 & 0.26797 & 0.013072 \\ 0 & 0.013072 & 0.63399 \end{pmatrix} \end{array}$$

Fig. 3 Transition Probability Matrix for Mandalay Station during May-Sept in 1998

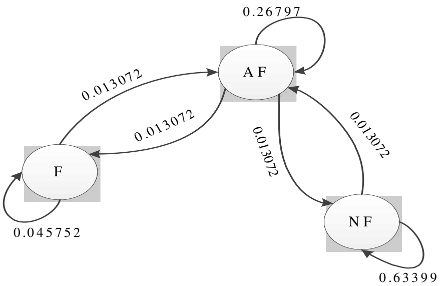


Fig. 4 Markov Model for Mandalay Station during May-Sept in 1998.

After getting the probability transition matrix, Markov state can be chosen by equal state (F, AF, and NF are 0.33) and flood state (F, AF, and NF are 1, 0 and 0 respectively). In this system, equal state is applied for river flood prediction.

5 Model Evaluation, Comparison, Error Results

In this section, the system evaluates water level of historical weather data from Mandalay station at Middle region of Ayeyarwady river in Myanmar using Markov model. In Markov model, water level is divided by Flood (F), Approximately Flood (AF) and Non Flood (NF) depended on danger water level from Hydro/Met station as shown in Table 3. Water level is more than and equal to danger water level assumes Flood. Water level is between the danger water level and 100cm far from danger water level assumes Approximately Flood. Non Flood means water level is less than 100cm far from the danger water level. Water level data is taken during May-September.

Table 3 Detail of Water Level Values for each Met\Hydro Station

No.	Station Name	Region	Danger Water Level	Water Level (Min-cm)	Water Level (Max-cm)
1	Myitkyinar	Upper	1200	219	1410
2	Sagaing	Middle	1150	197	1274
3	Mandalay	Middle	1260	414	1382
4	Magway	Middle	1700	437	1894
5	Aunglan	Lower	2550	1162	2737
6	Pyay	Lower	2900	1724	3025

In Markov model, there are two sections: daily forecast and yearly forecast. The system evaluated 3 days (daily forecast) for 16-18, June, 1999 at Mandalay station as an example. The evaluation results are shown in Fig. 5.

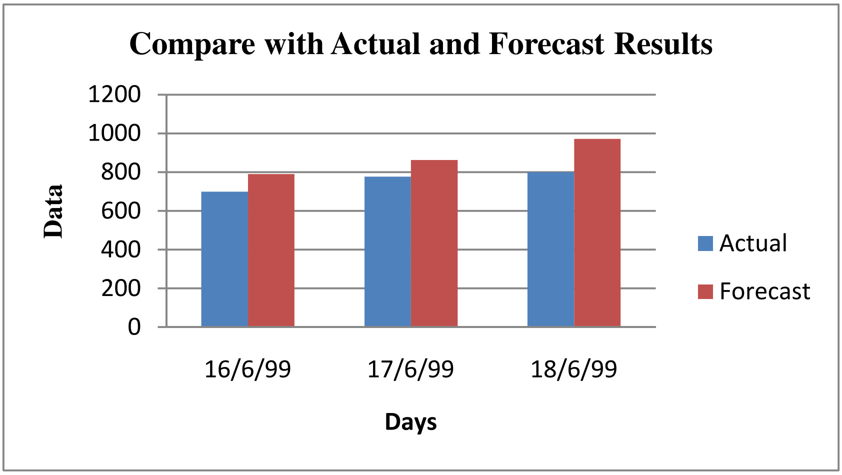


Fig. 5 Compare with Actual and Forecast Results for 3 days forecast at Mandalay Station

The system also compares the actual and forecast results, and the mean absolute percentage error (MAPE) are calculated between them as shown in Table 4.

Table 4 Actual, Forecast and Error Results for 3 days forecast at Mandalay Station

No.	Date	Actual	Forecast	MAPE
1	16/6/99	699	790	13.02 %
2	17/6/99	776	863	11.21 %
3	18/6/99	799	972	21.65 %

6 Conclusion

In this work, daily flood level prediction for Mandalay weather station is implemented by applying Markov model. It is characterized by real time of data flow, which is from the recording equipment in testing of complex systems. Weather station is selected by along with Ayeyarwady river which often face local flooding in rainy seasons as our region of interest. Flood level can be modeled by using various time scales such as yearly, monthly, weekly and daily. Hydrologic process such as daily or monthly water level may be well represented by stationary linear model such as Markov process. The proposed Markov model forecasted Weather data at Mandalay station by Daily to Daily. As an error rate of MAPE, the constructed model is useful to be acceptable accuracy for flood prediction of Ayeyarwady river in Myanmar. The system intends to extend and predict rainfall, snow melt, temperature and other economical effects.

References

1. Banik, S., Anwer, M., Khodadad Khan, A.F.M., Rouf, R.A., Chanchary, F.H.: Forecasting Bangladeshi monsoon rainfall using neural network and genetic algorithm approaches. International Technology Management Review **2**(1) (2009)
2. Hlaing, A.N.: Uncertain reasoning & prediction for river flooding with bayesian networks model. In: AICT (2010)
3. Kannan, M., Prabhakaran, S., Ramachandran, P.: Rainfall Forecasting Using Data Mining Technique. International Journal of Engineering and Technology **2**(6), 397–401 (2010)
4. Kulkarni, M.A., Patil, S., Rama, G.V., Sen, P.N.: Wind speed prediction using statistical regression and neural network. J. Earth Syst. Sci. **117**(4), 457–463 (2008)
5. Nayak, D., Mahapatra, A., Mishra, P.: A Survey on Rainfall Prediction using Artificial Neural Network. International Journal of Computer Applications **72**(16) (2013)
6. Nkrintra, S., et al.: Seasonal Forecasting of Thailand Summer Monsoon Rainfall. International Journal of Climatology **25**(5), 649–664 (2005). American Meteorological Society
7. Sen, N.: New forecast models for Indian south-west Monsoon season Rainfall. Current Science **84**(10), 1290–1291 (2003)

8. El-Shafie, A., Jaafer, O., Seyed, A.: Adaptive neuro-fuzzy inference system based model for rainfall forecasting in Klang River, Malaysia. International Journal of the Physical Sciences **6**(12), 2875–2888
9. Sohn, T., Lee, J.H., Lee, S.H., Ryu, C.S.: Statistical Prediction of Heavy Rain in South Korea. Advances in Atmospheric Sciences **22**(5), 703–710 (2005)
10. Song, H., Li, G.: Tourism demand modelling and forecasting - A review of recent research. Tourism Management **29**(2), 203–220 (2008)
11. Thein, U.H.M., Tun, U.S.: Floods, Forecasting System and Management in Myanmar. Ministry of Transport, Union of Myanmar

Multiplier Changing Overclocking Method for Processor Intensive Tasks

Win Win Aung

Abstract Overclocking is not a new method for most PC users with the most cost effective perspective for all processors. Two main overclocking methods: Front-Size Bus (FSB) and multiplier changing are used as the popular trends. When the processor need to increase the performance improvement for heavy workloads, overclocking method increases the clock speed cycles by changing multiplier values. In fact, overclocking is the best method to implement with trial and error process which is not warranted by the manufacturers but it is capable. The detail description of types of intensive workloads and multiplier changing overclocking results are shown. The testbed of the paper is based on the specification of Intel Core i7 3770 CPU@3.40 GHz with the core speed 1596.19 MHz as well as bus speed 99.76. The most commonly tools such as CPU-Z, RealTemp, CoreTemp, FFT-z, OCCT tool and Benchmark programs testings are used to validate the system evaluation.

Keywords Overclocking · CPU-Z · Intensive tasks · RealTemp · CoreTemp · FFT-z · OCCT · Benchmark

1 Introduction

Overclocking is the squeezing process of processing power out of the outdated computers and it is safely upgrading the performance of the processors. Traditionally, two main types of overclocking methods are used: called Front Size Bus (FSB) and multiplier changing. The Front Size Bus changes the whole portions of the processor including RAM, memory, GPU (Graphics Processing Unit) and CPU (Central Processing Unit). But the multiplier only changes the performance of the processor and so it is safer than the other methods. The overclocking method is used by the CPU-intensive tasks like video encoding, 3D rendering or gaming. But there is the main drawback which is shortening the life of the processor or permanently damages all of the components. Moreover, the stress testing programs and benchmarking are required when overclocking is performed [1].

W.W. Aung(✉)
University of Computer Studies, Mandalay, Myanmar
e-mail: nng.winwinaung@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/ 978-3-319-23204-1_45

Boosting the performance of a computer by obtaining the extra computational speed from the central processing unit with driving a computer system at higher clock frequencies in the default settings is the inexpensive process. But overclocking process varies depend on a device varies. However, the users can inevitably face the problems like excessive heat generation, quicker battery consumption, instability of the processor and potential hardware damage [2].

Overclocking is reasonable for high margin of the safety provided by the manufacturer, marketing reasons that can cause the manufacturer set lower default characteristics than possible or faster components usage than necessary. Most of all, the CPU frequency is much higher than the FSB frequency and it is generated by the multiplier [3].

The benchmarks can also detect the performance abilities of the current generation multiprocessing hardware such as multicore CPUs. Furthermore, benchmarking is used to accurate the robustness and weakness of the CPU. The CPU design influences the processing frequency in a computing system. In the modern CPUs, there are many factors included for versatile application. Most of the modern computers heavily rely on multiple execution units, multi-core at the hardware level and pipelined instructions to increase performance [4]. The subtopics included in the paper are as follows. Related work is in section 2, disadvantages and advantages of overclocking is in section 3, implementations and findings are in section 4 and the last section is conclusion and future work.

2 Overclocking's Related Work

Sasikal.D, M.Ravichandran M.E. and Dr.C.S.Ravichandran [5] presented dynamic overclocking mechanism which leads to appear to increase under longer duration of time and so this issue is solved with the help of throttling mechanism at some instances. Subramanian.V, Bezdek.M, Avirneni.N.D and Somani.A [6] showed various benchmark applications to overclock superscalar processors and the overheads of extra hardware and penalty of error recovery by adjusting the clock frequency dynamically. Hoffman.J.C and Pattichis.M.S [7] presented the voltage and temperature with normal processing conditions with the use of active feedback to maximize the overclocking preformance. Joo.B.G and Kim.S.M [8] approached to increase the speed of the device by overclocking and to optimize memory usage by deleting unnecessary apps by substituting more efficient ones. Perarnau.S and Huard.G [9] showed a tool to use in shared memory machines equipped with multiple CPU cores which are usually exploited concurrently by several users. Colwell.B [10] introduced better-than-worst-case design of overclocking. With normal worst-case design, any computing system is a conglomeration of components, operating within frequencies, power supply voltages, and temperature ranges.

3 Disadvantages and Advantages of Overclocking

The system is overclocked when the PC becomes obsolete and difficult to run the software without having sufficient performance. Overclocking the processors have simple objectives such as achieving flexible computation for different platforms, protecting the system's temperature under the proper conditions, emerging innovative users for future generations and reducing energy consumption for each cluster PCs by overclocking the centralized PC.

3.1 Disadvantages of Overclocking

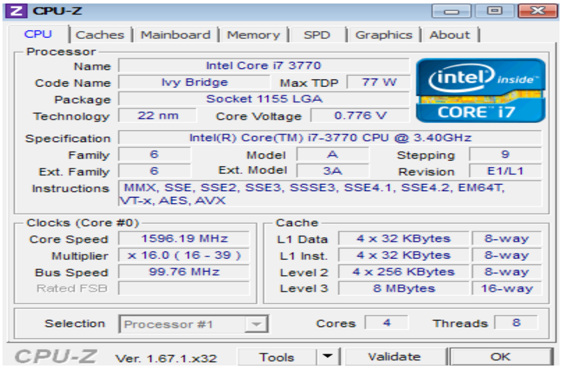
- It depends heavily on the cooling solutions.
- Need to monitor the system's temperature to exit between 45 and 60 degrees Celsius.
- Using programs in order to investigate the stability of the processor is another vital factor.
- Separating problems among Central Processing Unit (CPU), Random Access Memory (RAM) and Power Supply Unit (PSU) are sometimes very difficult to detect.
- Another important role is to achieve thermal balance and lifetime reliability of the system.

3.2 Advantages of Overclocking

- Upgrading existing obsolete PC into more powerful systems without having extra cost.
- Suitable method to use in developing countries where there is a limited budget.
- Almost zero labor cost at setting up the new upgrade to achieve the efficient performance.
- Provide green environment for future users with the least energy consumption.

4 Implementations and Findings

4.1 Capturing the Processor Specification Using CPU-Z Tool



Processor
Name Intel Core i7 3770
Code Name Ivy Bridge Max TDP 77 W
Package Socket 1155 LGA
Technology 22 nm Core Voltage 0.776 V
Specification Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz
Family 6 Model A Stepping 9
Ext. Family 6 Ext. Model 3A Revision E1/L1
Instructions MMX, SSE, SSE2, SSE3, SSSE3, SSE4.1, SSE4.2, EM64T, VT-x, AES, AVX

Clocks (Core #0)
Core Speed 1596.19 MHz
Multiplier x 16.0 (16 - 39)
Bus Speed 99.76 MHz
Rated FSB

Cache
L1 Data 4 x 32 KBytes 8-way
L1 Inst. 4 x 32 KBytes 8-way
Level 2 4 x 256 KBytes 8-way
Level 3 8 MBytes 16-way

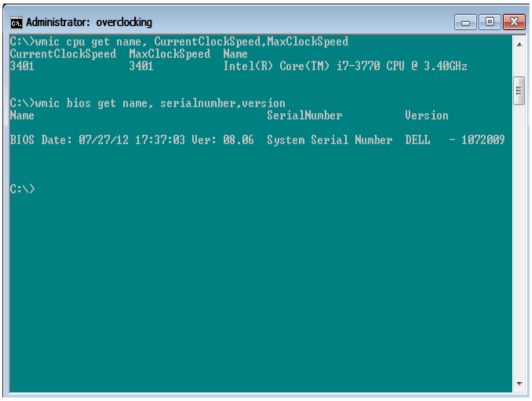
Selection Processor #1 Cores 4 Threads 8

CPU-Z Ver. 1.67.1.x32

Fig. 1 Capturing the specifications of the processor

According to the above CPU-Z tool detecting figure, the frequency of the core speed is 1596.19 MHz and the multiplier value is from 16 to maximum 39. The processor is Intel core i7 3770 with core voltage 0.776 volt and the bus speed is 99.76 MHz. When we multiply the bus speed (99.76) with the maximum multiplier value, we obtain the peak performance of the processor with 3890.64 GHz (nearly 4GHz).

4.2 Validating the CPU-Z Tool's Information with Command Prompt Detection

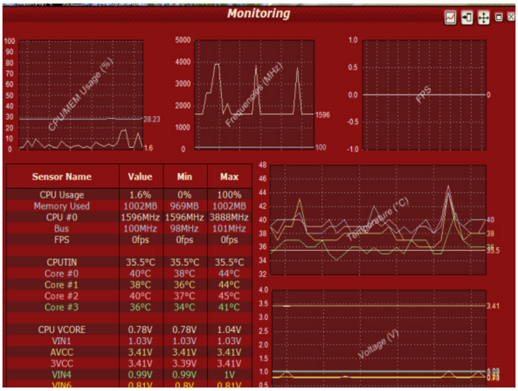


```
C:\>wmic cpu get name, CurrentClockSpeed,MaxClockSpeed
CurrentClockSpeed  MaxClockSpeed  Name
3401               3401           Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz


C:\>wmic bios get name, serialnumber,version
Name                                    SerialNumber          Version

BIOS Date: 07/27/12 17:37:03 Ver: 08.06  System Serial Number  DELL   - 1072009



C:\>
```

Fig. 2 Capturing the specifications of the processor via cprompt

The specification description from the tool is identical to the information provided by the cprompt including the name of the processor, the maximum clock speed and the current clock speed of the processor.

4.3 Overclocking Checking Tool Is Used to Upgrade the Processor Performance

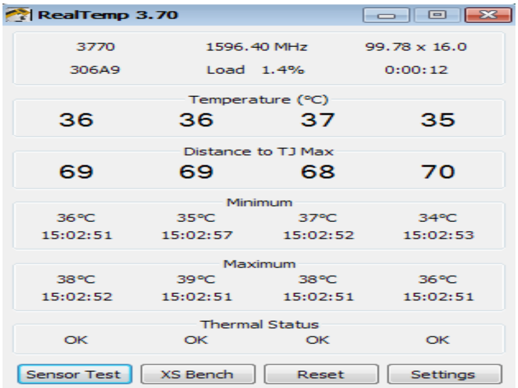


Sensor Name	Value	Min	Max
CPU Usage	1.6%	0%	100%
Memory Used	1002MB	969MB	1002MB
CPU #0	1596MHz	1596MHz	3888MHz
Bus	100MHz	98MHz	101MHz
FPS	0fps	0fps	0fps
CPUTIN	35.5°C	35.5°C	35.5°C
Core #0	40°C	38°C	44°C
Core #1	38°C	36°C	44°C
Core #2	40°C	37°C	45°C
Core #3	36°C	34°C	41°C
CPU VCORE	0.78V	0.78V	1.04V
VIN1	1.03V	1.03V	1.03V
AVCC	3.41V	3.41V	3.41V
3VCC	3.41V	3.39V	3.41V
VIN4	0.99V	0.99V	1V
VIN6	0.81V	0.8V	0.81V

Fig. 3 CPU and Memory usage of the original processor with OCCT

The above figure 3 shows the idle condition of the processor with current CPU speed 1596.19 MHz. And the CPU value is 1.6% from minimum 0% to maximum 100% and the memory condition is in the normal status with 1002 MB from 969 MB to 1002 MB. And also shows CPU temperature input and core voltage of the CPU.

4.4 The RealTemp Temperature Running before Overclocking


RealTemp 3.70

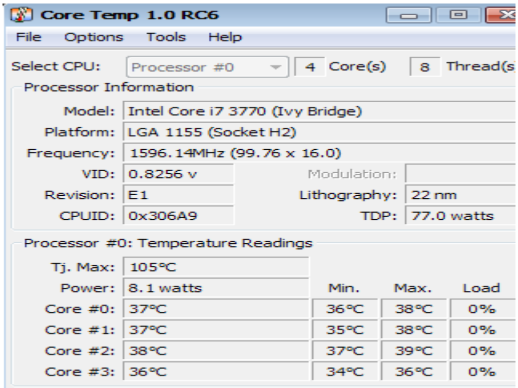
3770	1596.40 MHz	99.78 x 16.0
306A9	Load 1.4%	0:00:12

Temperature (°C)	36	36	37	35
Distance to TJ Max	69	69	68	70
Minimum	36°C 15:02:51	35°C 15:02:57	37°C 15:02:52	34°C 15:02:53
Maximum	38°C 15:02:52	39°C 15:02:51	38°C 15:02:51	36°C 15:02:51
Thermal Status	OK	OK	OK	OK

Fig. 4 The 4 individual core temperatures are presented

The temperature values of Core#0 are 36, Core#1 is 36, Core #2 is 37 and Core#3 is 35 respectively.

4.5 The CoreTemp Tool Is Used to Detect the Watts Usage and Individual Core Temperature


Core Temp 1.0 RC6

Select CPU: Processor #0 4 Core(s) 8 Thread(s

Processor Information

Model: Intel Core i7 3770 (Ivy Bridge)
Platform: LGA 1155 (Socket H2)
Frequency: 1596.14MHz (99.76 x 16.0)
VID: 0.8256 v
Modulation:
Revision: E1
Lithography: 22 nm
CPUID: 0x306A9
TDP: 77.0 watts

Processor #0: Temperature Readings

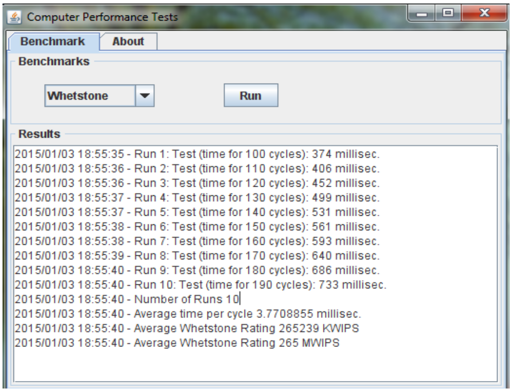
Tj. Max: 105°C

		Min.	Max.	Load
Power:	8.1 watts			
Core #0:	37°C	36°C	38°C	0%
Core #1:	37°C	35°C	38°C	0%
Core #2:	38°C	37°C	39°C	0%
Core #3:	36°C	34°C	36°C	0%

Fig. 5 The power usage and individual core information

Before overclocking the processor, the watts usage is only 8.1 watts and detection of the individual core values are little difference (for about 1 degree Celsius) from the RealTemp tool.

4.6 *Running Processor Intensive Tasks (Benchmark Programs)*



Benchmarks: Whetstone

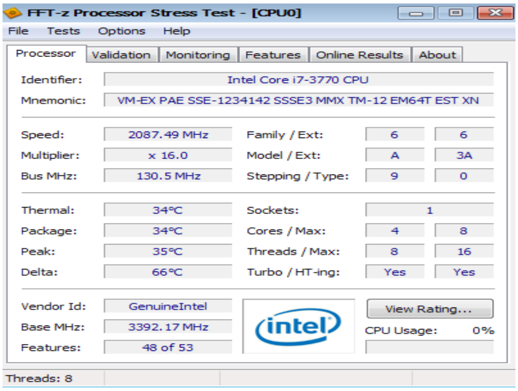
Results

```
2015/01/03 18:55:35 - Run 1: Test (time for 100 cycles): 374 millisec.
2015/01/03 18:55:36 - Run 2: Test (time for 110 cycles): 406 millisec.
2015/01/03 18:55:36 - Run 3: Test (time for 120 cycles): 452 millisec.
2015/01/03 18:55:37 - Run 4: Test (time for 130 cycles): 499 millisec.
2015/01/03 18:55:37 - Run 5: Test (time for 140 cycles): 531 millisec.
2015/01/03 18:55:38 - Run 6: Test (time for 150 cycles): 561 millisec.
2015/01/03 18:55:38 - Run 7: Test (time for 160 cycles): 593 millisec.
2015/01/03 18:55:39 - Run 8: Test (time for 170 cycles): 640 millisec.
2015/01/03 18:55:40 - Run 9: Test (time for 180 cycles): 686 millisec.
2015/01/03 18:55:40 - Run 10: Test (time for 190 cycles): 733 millisec.
2015/01/03 18:55:40 - Number of Runs 10
2015/01/03 18:55:40 - Average time per cycle 3.7708855 millisec.
2015/01/03 18:55:40 - Average Whetstone Rating 265239 KWIPS
2015/01/03 18:55:40 - Average Whetstone Rating 265 MWIPS
```

Fig. 6 Running the Benchmark before overclocking

According to the above figure, the number of 10 runs with average time per cycle 3.7708855 millisec. From run1 to run10 are (3.74, 3.69, 3.76, 3.83, 3.79, 3.74, 3.70, 3.76, 3.81, 3.85) respectively. And so it can gain 37.67 for totally which produces 3.76 milliseconds for 10 runs.

4.7 *FFT-z Tool Running to Detect the CPU Usage*

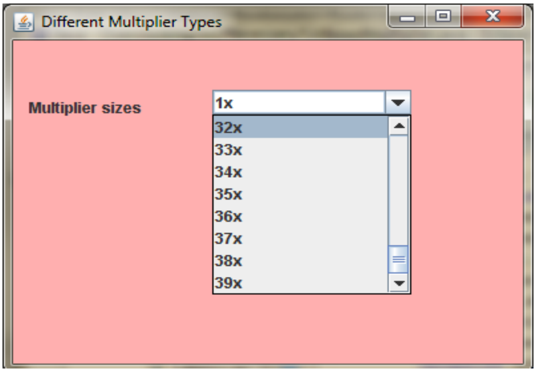


Identifier: Intel Core i7-3770 CPU
Mnemonic: VM-EX PAE SSE-1234142 SSSE3 MMX TM-12 EM64T EST XN
Speed: 2087.49 MHz
Multiplier: x 16.0
Bus MHz: 130.5 MHz
Family / Ext: 6 6
Model / Ext: A 3A
Stepping / Type: 9 0
Thermal: 34°C
Package: 34°C
Peak: 35°C
Delta: 66°C
Sockets: 1
Cores / Max: 4 8
Threads / Max: 8 16
Turbo / HT-ing: Yes Yes
Vendor Id: GenuineIntel
Base MHz: 3392.17 MHz
Features: 48 of 53
CPU Usage: 0%
Threads: 8

Fig. 7 The CPU usage capturing with FFT-z tool

The CPU only used zero percentage before overclocking the processor. The thermal condition is only 34·C and the peak value is only 35·C respectively.

4.8 Selecting the Multiplier Value


Different Multiplier Types

Multiplier sizes 1x

Fig. 8 Choosing the multiplier value

The value of multiplier is selected based on the specification of the processor. According to the Intel Core i7 3770 CPU@3.40 GHz, the reasonable value is selected to achieve the best performance for the processor intensive tasks.

4.9 Increasing the CPU Usage with Overclocking

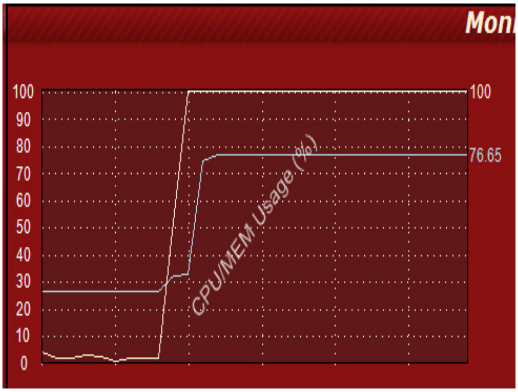


Fig. 9 The CPU usage increasing with multiplier value 38x

The value of the CPU becomes fully usage with the multiplier value 38x and the performance increased in 3790.88 MHz. The memory usage is 76.65.

4.10 Detecting the CPU Usage from Task Manager before Overclocking

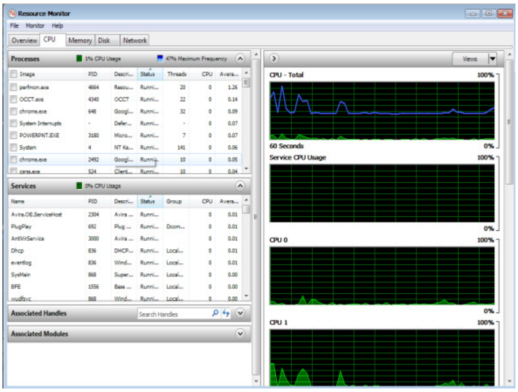

Resource Monitor
File Monitor Help
Overview CPU Memory Disk Network
Processes
1% CPU Usage
47% Maximum Frequency
Services
0% CPU Usage
Associated Handles
Search Handles
Associated Modules
Views
CPU - Total
100%
60 Seconds
0%
Service CPU Usage
100%
0%
CPU 0
100%
0%
CPU 1
100%

Image	PID	Descri...	Status	Threads	CPU	Avera...
☐ perfmon.exe	4664	Resou...	Runni...	20	0	1.26
☐ OCCT.exe	4340	OCCT	Runni...	22	0	0.14
☐ chrome.exe	648	Googl...	Runni...	32	0	0.09
☐ System Interrupts	-	Defer...	Runni...	-	0	0.07
☐ POWERPNT.EXE	3180	Micro...	Runni...	7	0	0.07
☐ System	4	NT Ke...	Runni...	141	0	0.06
☐ chrome.exe	2492	Googl...	Runni...	10	0	0.05
☐ csrss.exe	524	Client...	Runni...	10	0	0.04

Name	PID	Descri...	Status	Group	CPU	Avera...
Avira.OE.ServiceHost	2304	Avira ...	Runni...		0	0.01
PlugPlay	692	Plug ...	Runni...	Dcom...	0	0.01
AntiVirService	3000	Avira ...	Runni...		0	0.01
Dhcp	836	DHCP...	Runni...	Local...	0	0.01
eventlog	836	Wind...	Runni...	Local...	0	0.01
SysMain	868	Super...	Runni...	Local...	0	0.00
BFE	1556	Base ...	Runni...	Local...	0	0.00
wudfsvc	868	Wind...	Runni...	Local...	0	0.00

Fig. 10 Before overclocking condition of the processor

The value of the processor condition is still stable before overclocking.

4.11 After the Overclocking Conditions for the CPU

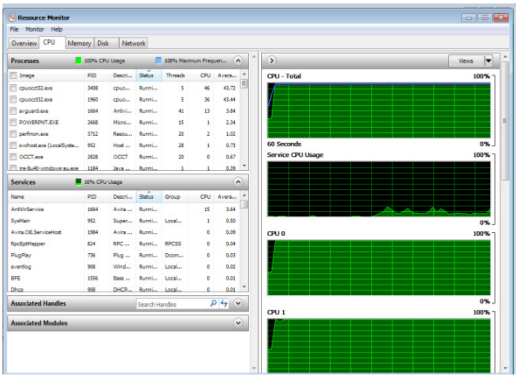

Resource Monitor
File Monitor Help
Overview CPU Memory Disk Network
Processes
100% CPU Usage
108% Maximum Frequen...
Services
16% CPU Usage
Associated Handles
Search Handles
Associated Modules
Views
CPU - Total
100%
60 Seconds
0%
Service CPU Usage
100%
0%
CPU 0
100%
0%
CPU 1
100%

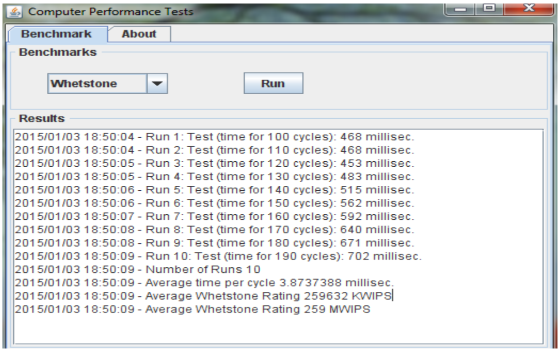
Image	PID	Descri...	Status	Threads	CPU	Avera...
☐ cpuocct32.exe	3408	cpuo...	Runni...	5	46	43.72
☐ cpuocct32.exe	1960	cpuo...	Runni...	5	36	43.44
☐ avguard.exe	1664	Antivi...	Runni...	41	13	3.84
☐ POWERPNT.EXE	2668	Micro...	Runni...	15	1	2.34
☐ perfmon.exe	3712	Resou...	Runni...	20	2	1.02
☐ svchost.exe (LocalSyste...	952	Host ...	Runni...	28	1	0.73
☐ OCCT.exe	2828	OCCT	Runni...	20	0	0.67
☐ jre-8u40-windows-au.exe	1184	Java ...	Runni...	1	1	0.39

Name	PID	Descri...	Status	Group	CPU	Avera...
AntiVirService	1664	Avira ...	Runni...		15	3.64
SysMain	952	Super...	Runni...	Local...	1	0.50
Avira.OE.ServiceHost	1984	Avira ...	Runni...		0	0.09
RpcEptMapper	824	RPC ...	Runni...	RPCSS	0	0.04
PlugPlay	736	Plug ...	Runni...	Dcom...	0	0.03
eventlog	908	Wind...	Runni...	Local...	0	0.02
BFE	1556	Base ...	Runni...	Local...	0	0.01
Dhcp	908	DHCP...	Runni...	Local...	0	0.01

Fig. 11 After overclocking condition of the processor

The value of the processor condition is increased significantly as a result of overclocking.

4.12 The Results of Benchmark Tasks after Overclocking

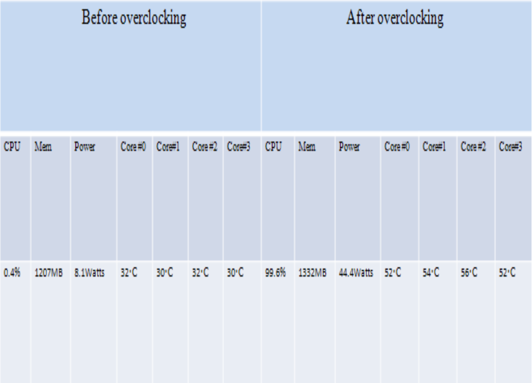


```
2015/01/03 18:50:04 - Run 1: Test (time for 100 cycles): 468 millisec.
2015/01/03 18:50:04 - Run 2: Test (time for 110 cycles): 468 millisec.
2015/01/03 18:50:05 - Run 3: Test (time for 120 cycles): 453 millisec.
2015/01/03 18:50:05 - Run 4: Test (time for 130 cycles): 483 millisec.
2015/01/03 18:50:06 - Run 5: Test (time for 140 cycles): 515 millisec.
2015/01/03 18:50:06 - Run 6: Test (time for 150 cycles): 562 millisec.
2015/01/03 18:50:07 - Run 7: Test (time for 160 cycles): 592 millisec.
2015/01/03 18:50:08 - Run 8: Test (time for 170 cycles): 640 millisec.
2015/01/03 18:50:08 - Run 9: Test (time for 180 cycles): 671 millisec.
2015/01/03 18:50:09 - Run 10: Test (time for 190 cycles): 702 millisec.
2015/01/03 18:50:09 - Number of Runs 10
2015/01/03 18:50:09 - Average time per cycle 3.8737388 millisec.
2015/01/03 18:50:09 - Average Whetstone Rating 259632 KWIPS
2015/01/03 18:50:09 - Average Whetstone Rating 259 MWIPS
```

Fig. 12 After overclocking condition of the benchmark

The benchmark results show the increasing from 3.77 millisecond to 3.87 millisecond respectively.

4.13 The Comparison Outcomes



Before overclocking							After overclocking						
CPU	Mem	Power	Core #0	Core#1	Core #2	Core#3	CPU	Mem	Power	Core #0	Core#1	Core #2	Core#3
0.4%	1207MB	8.1Watts	32◦C	30◦C	32◦C	30◦C	99.6%	1332MB	44.4Watts	52◦C	54◦C	56◦C	52◦C

Fig. 13 After overclocking condition of the benchmark

The benchmark conditions have been changed after overclocking the processor. The CPU value is increased for 9.5%. And also the memory is 125 MB, power is 36.3, Core #0 is 20°C, Core #1 is 24°C, Core #2 is 24°C and Core #3 is 22°C.

5 Conclusion and Future Work

Overclocking method is successfully implemented on Intel Core i7 3770 CPU@3.40 GHz with the core speed 1596.19 MHz and the bus speed is 99.76 respectively. When this method is applied, relevant tools can give greatly benefit to the system user and so there should be awareness of these useful tools. Finally the performance of the processor obtained nearly 3790.88 MHz and this is the best value for overclocking. Moreover, the thermal condition and watts usage of the processor is also shown individually. This indicates for future users to keep monitoring the CPU temperature not to exceed extremely. In coming ages, this method should be applied on the client server environment with the centralized one. Furthermore, in client side, the innovators should introduce thin clients in order to reduce, reuse and recycle for energy consumption processes which are the obstacles of green environments.

References

1. http://lifehacker.com/a-beginners-introduction-to-overclocking-your-intel-processor
2. Glen, G.: Newport High School, NP20 7YB UK. Investigating overclocking an Android mobile phone with a designed experiment (2012)
3. Ahmed, H., Azween, A.: IJCSNS International Journal of Computer Science and Network Security. The Impact of Overclocking the CPU to the Genetic Algorithm **9**(5) (May 2009)
4. Christopher, C., Christopher, W., Timothy, F.: Computing Performance Benchmarks among CPU, GPU, and FPGA (2012)
5. Sasikala, D., Ravichandran, M.E.M., Ravichandran, C.S.: Processor Performance Enhancement Using Self-Adaptive Clock Frequency. International Journal of Computer Applications (0975 – 8887) **3**(11) (July 2010)
6. Subramanian, V., Bezdek, M., Avirneni, N.D., Somani, A.: Superscalar Processor Performance Enhancement Through Reliable Dynamic Clock Frequency Tuning (2007)
7. Hoffman, J.C., Pattichis, M.S.: A High-Speed Dynamic Partial Reconfiguration Controller Using Direct Memory Access Through a Multiport Memory Controller and Overclocking with Active Feedback. International Journal of Reconfigurable Computing **2011**, Article ID 439072 (2011)
8. Joo, B.G., Kim, S.M: A User's Experience in Optimizing Smartphone Performance Using Overclocking and Memory Cleaning Techniques. International Journal of Smart Home **6**(4) (October 2012)
9. Swann, P., Guillaume, H.: KRASH: Reproducible CPU Load Generation on Many Cores Machines, Bangalore, India, January 9–14, 2001
10. Bob, C.: The Zen of Overclocking (March 2004)

Test Path Optimization Algorithm Compared with GA Based Approach

Aye Aye Kyaw and Myat Myat Min

Abstract Software testing performs a vital and basic role in developing software that is freedom from bugs and defects. Various levels of testing use Unified modeling language (UML) to design the tests. The test case generation from design specification is an essential work in testing phase. In recent trend, Model-Based testing approach attracts many researchers by using some search optimization techniques to produce an automated optimal test path. The best test path is got more beneficial than the other generated test paths because it can reduce time and cost more than others. In this paper, we compare and analysis different methodologies of model based best test path generation by using the UML activity diagram. We discuss the results the resulting test paths with respect to correctness and time consuming.

Keywords Model-Based testing · Test path · Test path generation · UML activity diagram

1 Introduction

A software project is made up of a series of phases. Generally, software projects comprise the phases as: i) Requirements gathering and analysis, ii) Design, iii) Development or Coding, iv) Testing, and vi) Deployment and Maintenance. Each and every phase has its own functionalities, milestones and deliverables to be achieved, in order to ensure a product that meet the needs of the customer.

This figure shows how defects from early phase add to the cost of the software project [1]. This implies that defects, when introduced in the requirements phase that might occur due to the misunderstanding or inconsistent requirements may inquire additional costs in detecting and correcting those errors. The defect can have continuously from requirement phase to testing phase. If the programmer knows having the defect at the testing phase, it raises the cost more and more at

---

A.A. Kyaw(✉) · M.M. Min
University of Computer Studies, Mandalay, Myanmar
e-mail: {ayeayekyaw2009,myatiimin}@gmail.com

© Springer International Publishing Switzerland 2016

T.T. Zin et al. (eds.), *Genetic and Evolutionary Computing*,
Advances in Intelligent Systems and Computing 387,
DOI: 10.1007/ 978-3-319-23204-1_46

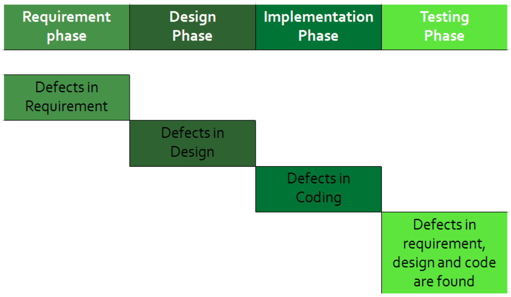


Fig. 1 Defects from early phase adding to the cost

every phase. To find defects early, testing activities shall be started as early as possible in the software or system development life cycle, and shall be focused on defined objectives. Good software testers cannot avoid models. MBT has emerged as a useful and efficient testing method for realizing adequate test coverage of systems.

Model based test case generation approaches identify faults in the implementation at early stage, reduce the software development time, and inspire developer to improve design quality. Unified Modeling Language (UML) has become the de facto standard for object oriented modeling and design. Most of researchers have used the following architecture for generating the test paths. There are various kinds of UML diagrams such as class diagram, activity diagram, sequence diagram, object diagram, and others. One or more UML diagram are used as an input, construct the dependency table based on these diagrams, and then create dependency graph according to the dependency table. Finally, produce all possible test paths. Some use data mining algorithms such as GA, Tabu Search, Ant Colony Optimization algorithm to find the best test path [2].

The rest of the paper is organized as follows: the related work is described in section 2. Test path optimization algorithm and GA-based algorithm to optimize the best test path are described in Section 3 and 4 respectively. Section 5 describes the experimental evaluation results with the different activity diagrams of the different domains. The paper is concluded with section 6.

2 Related Work

Many researchers have been successfully proposed test case generation for various software under many circumstances, using mainly search optimization techniques such as Genetic Algorithm, Ant Colony Optimization Algorithm, Tabu Search Algorithm, etc. The following papers are based work evolved under model based test case generation.

In [3], Shantghi and Mohan Kumar proposed an approach for generating test cases for object oriented software that uses Tabu Search technique to design and derive test cases. The Tabu search technique was applied to generate optimal test case.

Model Based Test Case Generation Technique Using Genetic Algorithms presented by M Sumalatha and Raju based on UML activity diagram. Finding all paths genetic algorithm operations are applied to obtain the best test case [4]. In case study of their paper, the expected results should be accurately. As the above generated new individuals are not part of the generated paths, Path 1 is the best node that is selected. By really Path 4 that is covered the whole of system should be the best path.

S. Srivastava, S. Kumar and A K Verma proposed a technique for optimal path sequencing in basis path testing. This approach is used to generate optimal paths using control flow graph (CFG) and Cyclomatic Complexity for finding the number of feasible paths. This paper applied ant colony optimization algorithm to prevent from selecting infeasible path and also prioritize the feasible paths i.e. which path is to be executed first to optimize time and complexity. With great ease of the ant colony optimization algorithm good results can be got and thus basis path testing can be improved, but there are areas where it can be extended such as if repetition of states or nodes can be controlled which have multiple paths. After execution of this algorithm, it automatically selects the best path sequence [5].

Shanthi and Mohan Kumar [6] proposed a new model based approach for automated test cases for UML activity diagrams using genetic algorithm. With the help of ADT test path are generated, by applying the GA most prioritized test case are generated. For each node, the number of incoming nodes and the number of outgoing nodes are calculated and then evaluate fitness values. Next, the initial test data is selected by randomly. Lastly, the Prioritized test path is generated. The generated test cases apply the branch coverage criteria and the Cyclomatic Complexity coverage.

3 Test Path Optimization Algorithm (TPOA)

The proposed system uses the activity diagram as an input for the automated algorithm of generating test paths. The Modelio Software has the option of exporting the UML diagram to XMI file. Fig. 2. shows the proposed algorithm for automatic test path generation. This system generates all possible test paths based on the extracted information from XMI file according to the SearchEdge function.

Input: XMI file for Activity Diagram (AD).
Output: all possible test paths and the best test path.

```
begin
    k:=0; j:=0; countJoinIn:=0; countForkOut:=0;
    countDecision:=0;
    totalDecision:=total decision node of the AD;
    $TP_{kj++}$ := source node of InitialEdge.
    myNode= target node of InitialEdge.
    SearchEdge( myNode, InitialEdge);
    OptimalTestPath().
end
```

Fig. 2 Algorithm for Automatic Best Test Path Generation

```
  SearchEdge(Node s, Edge ee)
   begin
      nodeType := type of s node;
      If (nodeType != FinalNode &&
                              (k<=totalDecision*2))
        For each edge $e_i$ Є E                    // i=1,2,…,n
          sNode := source node of edge $e_{i;}$
     tNode := target node of edge $e_{i;}$
     if (sNode== s )
            if (nodeType==Action ||
                               nodeType==Merge)
                    Add s to $TP_{kj++}$ ;
            else if (nodeType==Fork)
               countForkOut : = s.getCountNode();
                    countForkOut++;
               s.setCountNode(countForkOut);
               if (countForkOut ==1)
                          Add s to $TP_{kj++}$ ;
               endif
            else if (nodeType==Join)
               countJoinIn := s.getCountNode();
               countJoinIn++;
               s.setCountNode(countJoinIn);
               if (countJoinIn == s.getCountIn())
                          Add s to $TP_{kj++}$ ;
               endif
            else
               if(countDecision<=totalDecision*2)
                          countDecision++;
                      Add s to $TP_{kj++}$ ;
               endif
            endif
       SearchEdge(tNode, $e_i$);
     endif
        endfor
      else
          Add s to $TP_{kj++}$ ; k++;
     countDecision := 0;
     Set count for Fork node and Join node with 0;
      endif
end
```

Fig. 3 Algorithm of function *SearchEdge*

```
OptimalTestPath( )
begin
    maxControlNode:= total control node of $TP_0$;
    maxTotalNode  := total node of $TP_0$ ;
    BestPath  := $TP_0$ ;
    for  each test path $TP_i$ Є TP           // i=1,2,…n.
      curControlNode := total control node of $TP_i$ ;
      curTotalNode  := total node of $TP_i$  ;
      if (curControlNode > maxControlNode )
        if (curTotalNode>= maxTotalNode)
          BestPath := $TP_i$ ;
          maxControlNode  := curControlNode ;
          maxTotalNode  := curTotalNode ;
        else
          BestPath := $TP_i$ ;
          maxControlNode  := curControlNode ;
        endif
      endif
      if (curControlNode  == maxControlNode )
        if ( curTotalNode>= maxTotalNode)
          BestPath := $TP_i$ ;
          maxControlNode  := curControlNode ;
          maxTotalNode  := curTotalNode ;
        endif
      endif
   endfor
   Display BestPath as the optimal test path;
  end
```

Fig. 4 Algorithm of function *OptimalTestPath*

And then, the system optimizes the best test path among from these generated test paths by using the *OptimalTestPath* function. Finally, the best test path to be tested first can be got within a minimum amount of time [7].

4 GA-Based Algorithm to Optimize the Best Test Path

This section describes the GA based algorithm and process of the algorithm for optimizing the best test path that should be tested firstly. This approach is based on the system that V.Mary Sumalatha and Dr G.S.V.P.Raju proposed [4]. The GA based algorithm is as follows.

1. Activity Diagram is designed using the Modelio software and this software has the option of exporting the UML diagram to XMI file.
2. Extract the detailed information about the activity diagram from the .xmi file.

3. Assign the weights to all edges in the extracted information.
    a. Start the process with the first edge, assign value one.
    b. Increment the value of the next edge connecting to the target node and so on.
    c. If the node is a decision node, then firstly increment the value of the true way until the path terminates.
    d. And also assign the value for the false way until the path is complete.
4. Generate the different test paths from the initial node to the final node.
5. Calculate the fitness value for each paths
    a. For each path calculate the cost of the path, $X=\sum_{i=1} wi$.
    b. Fitness Function $F(X)=X*X$;
6. Evaluate the probability of the individual as $p(i)=F(X)/\sum F(X)$.
7. To choose the individuals from the initial population produce a new generation of solutions by picking from the existing solutions with a desire for solutions which are more suitable than another.
    a. Create a random value.
    b. Divide the probability range into bins which are sized according to the fitness value of the solution which they represent.
    c. Pick the individuals that will form the basis of the next generation by generating the random values and seeing which bin they fall into.
8. Randomly mutate the genes in the population.
9. Add the resulted individuals to the next generation and again calculate the fitness of this generation.
10. Repeat this process until all the paths have been covered or the minimum number of generations is reached.
11. Best test path generated.
12. End.

The processing steps of above algorithm are same with the TPOA approach until the generated all possible test paths. The different steps start at here.

-To evaluate the fitness of each member of the population. The calculated fitness values for each individual are

a. Decode the cost to binary value.
b. Evaluate the fitness according to $F(X)=X*X$.

- Divide the range into four bins, sized according to the relative fitness of the solution which they represent. By generating uniform random values and seeing which bin they fall into we pick the four strings that will form the basis of the next generation.

-Randomly number generator decides to mate all strings together.

-Reevaluate fitness of the new generation. This would be the first step in generating a new generation of solutions.

5 Experimental Results

This section describes the experimental evaluation results in the TPOA and GA based algorithms. An evaluation approach is built to evaluate the TPOA model and to compare its performance with the GA based model. It compares their performance while generating test paths for an activity diagram taken from the different systems that differ in size and domain according to some factors described as follows:

-The total time taken of the execution process in both models.
-The correctness of the generated of the best test path for both models.

The comparison graphs for those factors depict the improvement in performance as well as the high quality of the test paths generated by the TPOA system.

A. The total time taken of the execution process in both models
The comparison graph describes the execution time with nano-seconds for all tested activity diagrams as shown in Fig. 5. The red color is given for our TPOA model while the green color is for the other. The execution time by GA increases as many as the total generated test paths of activity diagram. TPOA is not a significant different of the execution time between all activity diagram. The execution time of TPOA slightly increases at the highest total number of generated test paths in an activity diagram.

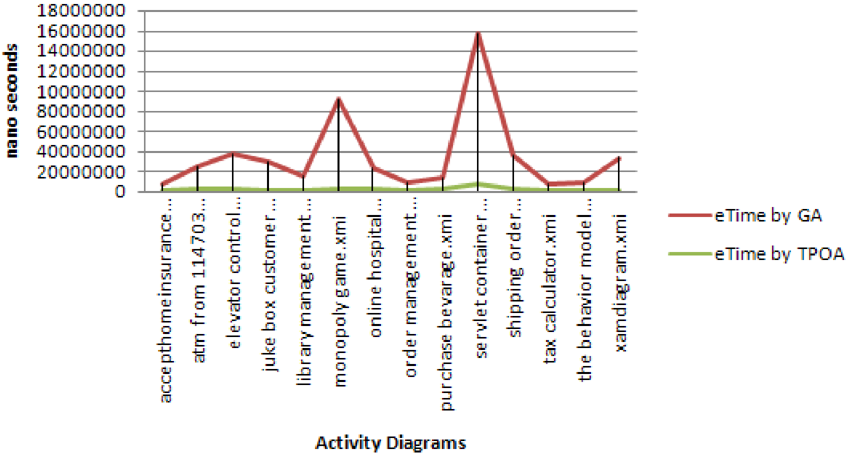


Fig. 5 Comparison of Execution Time

B. The correctness of the generated of the best test path for both models
Only one time has been utilized to generate the optimal test path within the TPOA model whereas in the other model it takes large amount of time required for optimizing the test path because of the usage of random value in GA based approach.

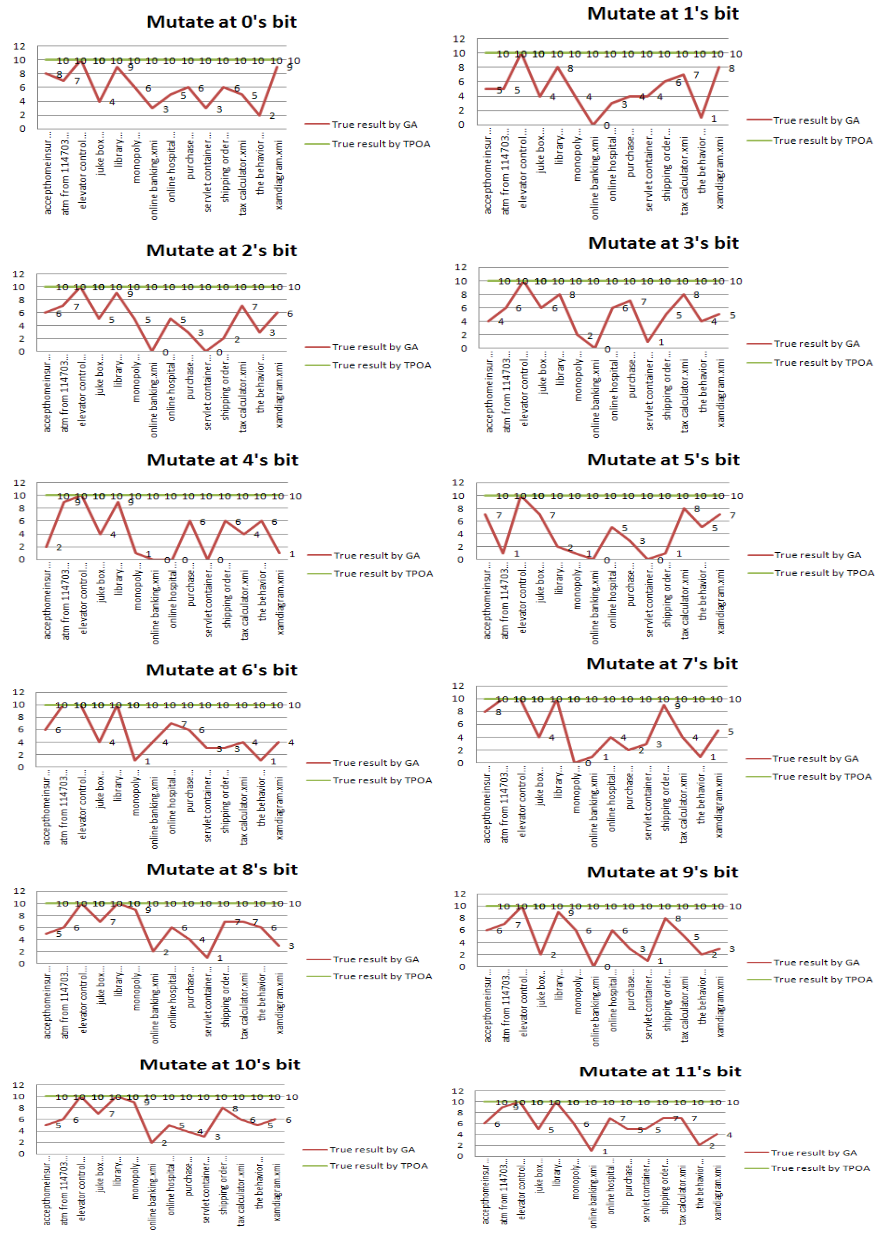


Fig. 6 Correctness of the generated best test path

In Fig. 6, the following line graphs compare the numbers of counted for getting the best test path per ten times on different activity diagrams by using the GA based and TPOA approach. Each line graph describes the results that get from the GA based approach, when the mutation point of each population changed, and the TPOA algorithm per times. Although the TPOA algorithm gets the same result with respect to this diagram for each activity diagram at all tested times, the GA based approach cannot result the same best path because it uses the random value for each generation. Each chromosome that is decoded from the cost of each test path has 12 bit binary value. In spite of changing the mutation point, the same result of each activity diagram cannot be got for all tested times.

6 Conclusion

To sum up, the TPOA system based on MBT develops more accurate all possible test paths due to generate directly from XMI file instead of using the dependency table and dependency graph. The TPOA system saves time in choosing the best test path because other optimization algorithms as the GA based approach spend time to calculate weight but the proposed approach does not need to evaluate weight. Therefore, the system used TPOA approach is more efficient and effective system than the GA based approach, because it reduces the human effort in finding bugs and errors, and also the steps at the evolution of generated test paths. The TPOA model saves time and effort. Besides, it also increases the quality of generated test cases. The TPOA approach will reduce the cost of software development and improve quality of the software.

References

1. http://www.softwareresults.us/defects-in-software.html
2. Kyaw, A.A., Min, M.M.: Model-based automatic optimal test path generation via search optimization techniques: a critical review. In: Proceedings of the 12th International Conference on Computer Applications (2014)
3. Shanthi, A.V.K., Mohan Kumar, G.: A Novel Approach for Automated Test Path Generation using TABU Search Algorithm. International Journal of Computer Applications (0975 – 888) **48**(13) (June 2012)
4. Mary Sumalatha, V., Raju, G.S.V.P: An Model Based Test Case Generation Technique Using Genetic Algorithms. The International Journal of Computer Science & Applications (TIJCSA) **1**(9) (November 2012). ISSN – 2278-1080
5. Srivastava, S., Kumar, S., Verma, A.K.: Optimal Path Sequencing in Basis Path Testing. International Journal of Advanced Computational Engineering and Networking **1**(1) (2013). ISSN (PRINT): 2320-2106
6. Shanthi, A.V.K., Mohan Kumar, G.: A Heuristic Technique for Automated Test Cases Generation from UML Activity Diagram. Journal of Computer Science and Applications. **4**(2), 75-86 (2012)
7. Kyaw, A.A., Min, M.M.: A proposed test path generation algorithm based on UML activity diagram. In: Proceedings of the 13th International Conference on Computer Applications (2015)

Author Index

Ashizawa, Yoshito 383
Aung, Swe Swe 327
Aung, Than Nwe 175
Aung, Win Win 445
Aye, Nyein 259

Cai, Sijing 3
Chang, Chin-Ie 247
Chen, Chia-Mei 319
Cui, Qiao-Chu 55
Cui, Wei 11

Fukuyama, Kiichi 429

Gagarina, L.G. 229
Geng, Guanggang 299

Hagita, Naoki 71
Hama, Hiromitsu 29, 65, 105, 361
Hasebe, Naoya 165
He, Yao 21
Hein, Chit La Pyae Myo 113
Hironaga, Mikiya 165
Hlaing, Yu Wai 189
Hnin, Than Thida 237
Hsu, Sung-Chien 319
Htay, Kyawt Kyawt 259
Huang, Yea-Shuan 247

Istanda, Vaci 3

Jiang, Shou-Da 45, 55

Kamada, Hiroshi 219
Kamolphiwong, Sinchai 289
Khaing, Phyu Phyu 175
Khin, Mie Mie 269, 435
Kobayashi, Ikuo 429
Kogi, Kazunobu 105
Ku, Ching-Heng 279, 299
Kuo, Sheng-Wei 279
Kyaw, Aye Aye 455

Lai, Gu-Hsin 319
Lee, Xiaodong 299
Li, Jun-Bao 11
Li, Yi-Fan 11
Liao, Lyu-Chao 3
Lin, Kyaw Zin 125
Lin, Lian-Lei 55
Lin, Ting-Wei 311
Liu, Xiaowei 299
Liu, Yuan 21
Liu, Zhongkan 29, 65
Lu, Ai-Chin 279
Lu, Xin 35
Lynn, Khin Thidar 237

Masuda, Kazuaki 219
Min, Myat Myat 455
Mitsui, Tatsuaki 411
Mon, Myat Thida 209
Morisaka, Tadamichi 105
Myo, Khin Mar 113
Myo, Zin Mar 209

Naing, Thinn Thu 155, 327
Nakagawa, Katsuji 383
Nakagawa, Yuki 429
Nanig, Thinn Thu 155
Nassary, Noor Agha 429
Noopetch, Wuttipon 289

Ohnuki, Shinichiro 383
Oo, Ko Ko 125
Oo, Kyaw May 189

Pan, Jeng-Shyang 3, 11, 21
Phway, Linn Linn 349
Phyo, Zar Lwin 369
Portnov, E.M. 125

Razak, Angger Abdul 403

Sakai, Mai 105
Sakakibara, Shota 421
San, Thinn Htet Htet 435
Sein, Myint Myint 339, 349

Su, Jun-Ming 311
Su, Qiang 11

Takahashi, Hideya 199
Takano, Yuta 383
Takubo, Tomohito 71
Tatsuzawa, Keisuke 383
Thwin, Mie Mie Su 137
Tin, Mie Mie 269
Tin, Pyke 361
Toriu, Takashi 165
Tsai, Geng-Da 279
Tsai, Pei-Wei 3, 21
Tseng, Shian-Shyong 279, 299, 311
Tun, Aung Kyaw 229
Tun, Hein 125

Ueno, Atsushi 71

Wang, Deng-Yi 95
Wang, Wen-Fong 83, 95
Watanabe, Koki 391
Wei, Chang-An 45, 55
Wutyi, Khaing Shwe 137

Yamada, Keisuke 219
Yamada, Kenji 199
Yan, Zhiwei 299
Yang, Ching-Yu 83, 95
Ye, Kyaw Zaw 125, 229
Yokota, Mitsuhiro 403, 411, 421
Yoshida, Shigeto 199
Yoshimoto, Kayo 199

Zar, Myat Thu 339
Zhang, Shunmiao 21
Zhao, Jia-Yu 45
Zhou, Mingyong 29, 35, 65
Zin, Thi Thi 361